DNA Replication in the human Major Histocompatibility Complex

Petros Takousis

A thesis submitted for the degree of
Doctor of Philosophy
at the
University of London

2007

Human Cytogenetics Laboratory
Cancer Research UK London Research Institute
44 Lincoln's Inn Fields
London WC2A 3PX

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration

I, Petros Takousis, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Copyright statement

Acknowledgements

I thank my supervisor Denise Sheer for giving me the opportunity to test and be tested, during the course of this study. I thank her for her advice and guidance throughout this period as well as her immense support with writing-up.

Many thanks to Rossi Christova, Rossen Donev, Alistair Newall and Diego Ottaviani for offering their help and expertise. I also thank other past and present members of the lab for their help, encouragement and useful discussions, or for having a pint with me after work. To Andi Bolzer, Paul Mulholland, Jayson Wang, Tim Forshew, Tania Jones, Radost Vatcheva, Pei Jun Wu, Ruth Tatevossian, Chiara Mazzanti, thank you very much.

It would be impossible to take this project further without the expert help of several other people. I deeply thank: Gary Warnes and Ayad Eddaoudi for help with flow cytometry; Veronique Azuara for help with the "elusive" BrdU immunoprecipitation; Alastair Nicol and Peter Jordan for help with confocal microscopy; everyone in the cell services team; Graham Clark and members of the Equipment Park; Richard Mitter, Mike Mitchell and Gavin Kelly for all their help with statistics and bioinformatics; Stephan Beck and his team for sharing their MHC arrays with us; Ian Giddings for help with array scanning and advice; Yiannis Chronakis for help with PCs; Stathis Sideris for crucial help with figures and image processing as well as other computer related issues; and last but not least, Ken Taylor for hardware and software support and lots of encouragement.

I have come to be very close friends with several of the people I have met over the last few years, who are either directly related to my project or whom I have simply met through work-related social occasions. I would like to thank them and dedicate this thesis to them as well as my close friends from before.

Ένα μεγάλο ευχαριστώ σ'όλη μου την οικογένεια που με βοήθησε, με στήριξε και εξακολουθεί να με στηρίζει. Χωρίς αυτούς δε θα μπορούσα να τα καταφέρω. Αφιερώνω την εργασία αυτή στην οικογένειά μου και τους στενούς μου φίλους που μου συμπαραστέκοντε. Σας ευχαριστώ!

Abstract

DNA replication is a vital component of the eukaryotic cell cycle. During the course of S-phase, numerous origins of replication become activated along each chromosome. Several adjacent origins fire synchronously to replicate large sections of a chromosome at specific times. Early studies identified a relationship between cytogenetic bands and replication timing: GC-rich R-bands replicate early while AT-rich and gene poor G-bands replicate late, apparently regardless of differentiation and developmental status. Subsequent studies revealed that other factors such as transcriptional status also influence the replication programme.

The aim of this thesis is to examine the organisation of DNA replication in the human Major Histocompatibility Complex (MHC) on chromosome 6, and understand how it relates to gene expression and inherent genomic properties. A previous investigation from the Human Cytogenetics Laboratory using fluorescence in situ hybridisation (FISH) suggested that replication timing of the MHC is organised into distinct zones, with the MHC class II region, an AT-rich isochore, replicating later than neighbouring regions.

Using a biochemical approach, the entire MHC was found to replicate within the first half of S-phase in cell lines derived from different tissues. Subsequent analysis of a B-lymphoblastoid cell line using a high resolution tiling path array for the MHC confirmed that a large proportion of the MHC class II replicates later than its neighbours. The data suggested the existence within the MHC of replication origins that fire at distinct times in S-phase.

An investigation of replication initiation in the MHC revealed the presence of several potential initiation sites, which were further analysed by quantitative PCR. The gene-rich MHC class III was found to have a relatively large number of replication initiation sites. Overall, these results suggest that either specific origins of replication or zones of initiation can fulfill the replication requirements of a region.

Table of Contents

List of tables and figures

Abbreviations

°C	degree Celsius
3D	three dimensional
aa	amino acid(s)
APC	antigen presenting cell
ATCC	American Type Culture Collection
ATP	adenosine triphosphate
ATPase	adenosine triphosphatase
BAC	bacterial artificial chromosome
bp	base pair(s)
BrdU	bromodeoxyuridine
BSA	bovine serum albumin
cdc	cell division cycle
cdk	cyclin dependent kinase
cDNA	complementary DNA
CHO	Chinese hamster ovary
CIITA	class II transactivator
CLIP	class II associated invariant chain peptide
CT	chromosome territory
DAPI	4'6'-diamino-2-phenylindole
DDK	Dbf4 dependent kinase
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
Dnase	deoxyribonuclease
dNTP	deoxy nucleotide triphosphate
DTT	dithiothreitol
EBV	Epstein Barr virus
ECACC	European Collection of Cell Cultures
EDTA	ethylenediaminetetracetic acid
ER	endoplasmic reticulum
FACS	fluorescence activated cell sorting
FCS	fetal calf serum
FISH	fluorescence in situ hybridisation
FITC	fluorescein isothiocyanate
GAS	gamma activated sequence
G-band	Giemsa-band
HAT	histone acetyltransferase
HDAC	histone deacetylase
HLA	human leukocyte antigen
HMGB/N	high motility group B/N
HP1	heterochromatin protein 1
HRE	hormone response element
HSP	heat shock protein
IAA	isoamyl alcohol
ICD	interchromatin domain
IFN	interferon

IFNGR	interferon gamma receptor
IP	immunoprecipitation
IRF	interferon response factor
Jak	Janus kinase
Kb	kilobase(s)
KDa	kilo-Dalton
LCR	locus control region
LD	linkage disequilibrium
LH	linker histones
LINE	long interspersed nuclear element
MAC	membrane attack complex
MAR/SAR	matrix/scaffold attachment region
Mb	megabase(s)
Mcm2-7	mini chromosome maintainance proteins 2-7
MHC	Major Histocompatibility Complex
mRNA	messenger RNA
MRS	MAR/SAR recognition sequence
mt DNA	mitochondrial DNA
NK	natural killer
NP-40	Nonidet P-40
ODP	origin decision point
ORC	origin recognition complex
ORF	open reading frame
PABL	pseudoautosomal boundary-like
PAR	pseudoautosomal region
PBS	phosphate buffered saline
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PI	propidium iodide
PLG	phase lock gel
Pol	polymerase
pRb	retinoblastoma protein
PSMB	proteasome beta subunit
RAR	retinoic acid receptor
R-band	Reverse-band
RC	replicative complex
rDNA	ribosomal DNA
RFC	replication factor C
RNA	ribonucleic acid
RNase	ribonuclease
RPA	replication protein A
rpm	revolutions per minute
RPMI-1640	Roswell Park Memorial Institute 1640
RXR	retinoid X receptor
SCF	Skp2-Cullin-F-box
SDS	sodium dodecyl sulphate
SINE	short interspersed nuclear element
SSB	single strand DNA binding protein
SSC	sodium chloride sodium citrate
ssDNA	single stranded DNA

Stat	signal transducer and activator of transcription
TAE	tris acetate EDTA
TAP	transporter associated with antigen processing
TAPBP	tapasin binding protein
TE	Tris EDTA
tiff	tagged image file
TNF	tumour necrosis factor
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
TSA	Trichostatin A
UCSC	University of California Santa Cruz
UV	ultraviolet
XIC	X inactivation centre

Chapter 1

Introduction

The formation of two cells from one by progression through the cell cycle is fundamental to the continuity of life. During synthesis- or S-phase during the cell cycle, the chromosomes reproduce themselves by a precise and orderly programme of DNA replication, so that the genetic make-up of the two daughter cells is identical (Figure 1.1). The aim of this thesis is to understand the organisation of DNA replication in a model genomic region, the human Major Histocompatibility Complex (MHC).

1.1 Organisation of the human genome

The human genome consists of approximately 3.08 billion base pairs. It encodes approximately 25000 genes (Collins, Lander et al. 2004), which reside in the 46 chromosomes: the 22 pairs of autosomes and the 2 sex chromosomes. Human chromosomes vary in size; chromosome 1 is the longest and chromosome 21 the shortest containing about 245 Mb and 27 Mb respectively. The number of genes on each chromosome varies, but does not show a strict correlation with chromosome length. Furthermore, regions across chromosomes have greatly variable gene density. Similarly, certain chromosomal regions have particular DNA sequence characteristics, such as high or low content in GC base pairs.

1.1.1 Chromosome architecture

Human chromosomes undergo major structural changes during the cell cycle. From G1 they proceed to become replicated during S-phase, so that each chromosome in the G2 phase consists of two identical sister chromatids. The chromosomes are then subjected to an extreme degree of condensation in preparation for mitosis. By the metaphase stage of mitosis, they adopt the familiar X shape, which represents the two

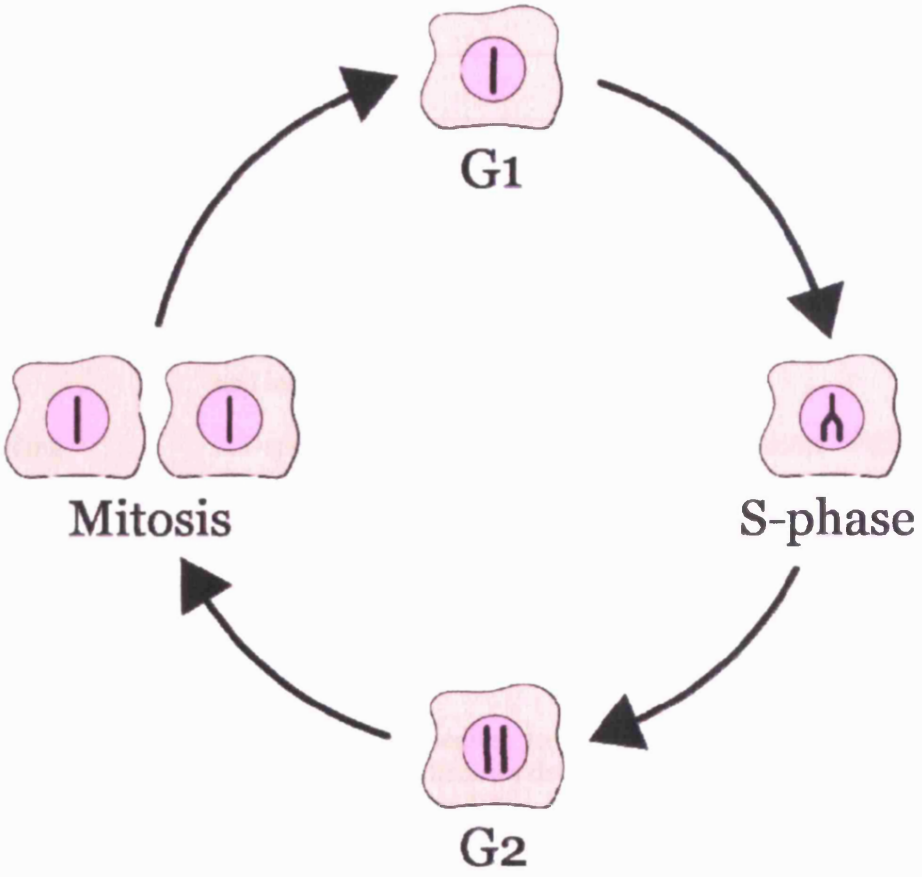


Figure 1.1 The cell cycle.

During S-phase, DNA synthesis takes places and by G2 chromosomes have become duplicated. G2 is followed by Mitosis when chromosomes separate, and after the mother cell divides, the two daughter cells are back to G1.

sister chromatids held together at the centromere. In this highly condensed stage, the 22 pairs of autosomes and 2 sex chromosomes can be classified according to their size and position of the centromere.

Banding	Description
G-banding	Giemsa staining after incubation with trypsin or salt solution at 60 °C
R-banding	Giemsa staining after incubation with hot (80-90 °C) alkaline or acidic saline (resulting in pattern reverse to G-banding)
C-banding	Giemsa staining of heterochromatin after acid/alkali treatment and incubation in hot SSC
Q-banding	AT-specific fluorochromes: quinacrine, Hoechst 33258 or DAPI (similar to G-banding)
Negative Q-banding	GC-specific fluorochromes: chromomycin, actinomycin D and mithramycin (similar to R-banding)
In situ hybridisation banding	Hybridisation of DNA repeat sequences to produce patterns similar to G- or R-bands
Antibody or immuno -cytogenetic banding	Binding of antibodies to specific chromatin or protein structures (e.g. antibodies detecting Z-DNA, 5-methylcytocine or acetylated histone H3)

Table 1.1 Chromosome banding techniques

Chromosomes at metaphase have characteristic chromosome bands that are seen using various treatments (Table 1.1). One of the most common methods of chromosome banding involves treating with trypsin followed by staining with Giemsa. Regions of a chromosome that are stained under these conditions are termed G bands as opposed to the Reverse of Giemsa or R bands that stain under different conditions (see Table 1.1). The pattern of bands on each chromosome is reproducible throughout cell development and differentiation. It is also unique and therefore facilitates chromosome identification (Figure 1.2). G- and R banding has been attributed to differences in nucleotide composition between the chromosomal regions. G bands are typical of AT-rich regions, in contrast to R bands which are GC-rich. Banding

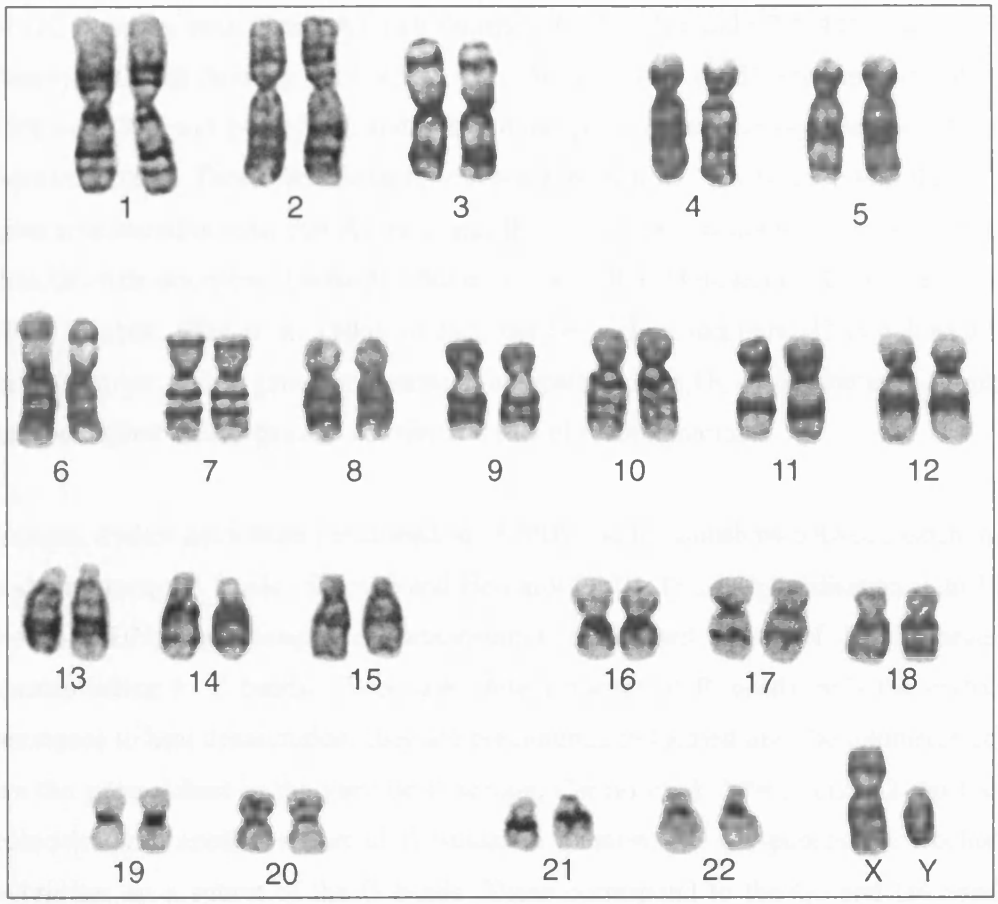


Figure 1.2 Normal human male karyotype with G-banding (courtesy of Dr Denise Sheer).

patterns similar to G- and R bands are produced by other staining procedures that have a propensity for AT- or GC-rich DNA (Verma and Babu 1995). Of significance for this thesis, the G- and R bands correlate with genomic regions that replicate late and early during S-phase, respectively, as discussed later.

1.1.2 Isochores

The human genome contains long DNA segments of approximately 300 Kb and above that have uniform sequence composition. These segments can be separated by ultra-centrifugation in density gradients of caesium salts (CsCl or Cs_2SO_4) as pioneered in the 1970s by Bernardi, since GC-rich DNA is denser than AT-rich (Macaya, Thiery et al. 1976; Cuny, Soriano et al. 1981). Five main classes have been identified in order

of GC richness: two "light" AT-rich families: L1 (~38%) and L2 (~40%), and three "heavy" GC-rich families: H1 (~45%), H2 (~50%) and H3 (~53%) representing about 30% and 33%, and 24%, 7.5% and 4-5% of the genome respectively (Bernardi 1995; Bernardi 2000). These five classes, which are known as "isochores", also differ in gene concentration such that AT-rich, and thus GC-poor, isochores have fewer genes than GC-rich isochores (Bernardi, Olofsson et al. 1985; Mouchiroud, D'Onofrio et al. 1991; Zoubak, Clay et al. 1996). In fact, the GC-richest isochore H3 is believed to have 20 times higher gene concentration compared to the GC-poor isochores. It also has the highest transcriptional activity and rate of recombination.

Various studies have been performed to identify the relationships between isochores and chromosomal bands (Saccone and Bernardi 2001). *In situ* hybridisation with H3 isochore DNA on metaphase chromosomes recognised a set of bands largely corresponding to T bands. These constitute a subset of R bands with the highest resistance to heat denaturation; they are predominantly located near the telomeres and are the gene richest in the genome (Saccone, Caccio et al. 1996). The H2 isochore coincides with another subset of R bands. In contrast, the GC-poorest L1 isochore hybridises on a subset of the G bands. These correspond to the G3 and G4 bands described by Francke, based on the intensity of their staining on metaphase chromosomes (Francke 1994). The observation that L1 DNA is absent from T bands and H3 DNA is absent from G3 and G4 bands, suggests that the L1 and H3 isochore families identify chromosomal regions with distinct characteristics. However, a clear banding pattern has not been observed with the H1 and L2 isochore families. These have been found on all the remaining R and G bands that had not been recognised by the H3 and L1 isochores respectively. This emphasises the proximity in GC content between fractions of these two sets of bands, which in cytogenetic terms have different characteristics. One example attesting to this "inconsistency" is that of the G bands 21q11.2 and 21q21.2 which have a higher GC level than the R bands 21q22.12 and 21q22.2 (Hattori, Fujiyama et al. 2000). This finding suggests that the fundamental difference between R and G bands is not merely due to differences in base composition but also the structure of the chromatin.

1.1.3 Repetitive sequences

It has been estimated that less than 2% of the human genome contains single-copy, gene coding sequences (Eyre-Walker and Hurst 2001), while a similar proportion is thought to be involved in gene regulation. Non-coding and non-regulatory DNA acquired the term "junk" DNA as it had no apparent function (Ohno 1972). This early characterisation is not favoured as much nowadays because some of these "junk" sequences may have functions that are not yet clearly understood, and because others are believed to have been conserved over millions of years of evolution (Dermitzakis, Reymond et al. 2005; Taylor 2005). Some of the non-genic parts of the genome are composed of sequences that are present with more than one copy in the genome. Two principal types of repeats can be found in the human genome: 1) the interspersed and 2) the simple sequence repeats.

Interspersed repeats form a large part of the human genome and are subdivided into two main categories: (i) the Short Interspersed Nuclear Elements, and (ii) the Long Interspersed Nuclear Elements, also referred to as SINEs and LINEs respectively. They propagate by retrotransposition, a process of "copy and paste" where the sequence of the repeat firstly becomes transcribed by RNA polymerase II or III, followed by reverse transcription and integration of the new copy into an available genomic site (Dewannieux and Heidmann 2005; Han and Boeke 2005).

SINEs have a relatively high GC content (56%) with the *Alu* elements being the dominant members. These have a consensus sequence which is about 300 bp long and as many as a million copies have been reported to exist within the genome (Lander 01) making up ~11% of the total. They are often located in the introns of genes as well as in the regions between genes and are suggested to have a role in alternative splicing (Kreahling and Graveley 2004).

The most abundant LINE family is *LINE-1* or *L1* (not to be confused with the L1 isochore). This is relatively GC poor (42%) compared to the *Alu* elements but has a longer consensus sequence of about 6 Kb and contains two open reading frames. ORF1 codes for an RNA binding protein and ORF2 encodes a protein with

endonuclease and reverse transcriptase activities. These allow *L1* elements to retrotranspose autonomously, unlike the *Alu* elements which usurp the L1 machinery in order to do so (Han and Boeke 2005). Over half a million copies of *L1* elements constitute ~17% of the human genome (Lander, Linton et al. 2001). Of those, fewer than hundred still have functional genes that can be transcribed and translated.

It is noteworthy that LINEs occur at higher density in AT-rich G bands, in antithesis to the SINEs which have a tendency to populate GC-rich R bands (Korenberg and Rykowski 1988). Both the SINE and LINE families of repeats, together with the smaller groups of LTR retrotransposons and DNA transposons, make up an astonishing 45% of the genome (Lander, Linton et al. 2001).

Simple sequence repeats consist of perfect or slightly imperfect tandem repeats. Tandemly repetitive sequences, commonly known as "satellite DNAs" are classified into three major groups: (i) the very highly repetitive "satellites" – organised as large clusters of up to 100 Mb, (ii) the moderately repetitive "minisatellites" – organised as 15 bp repeats up to 30 Kb, and (iii) "microsatellites" – organised as short 2-6 bp repeats with 10-100 copies. The term "satellite DNA" is assigned to DNA sequences forming a separate fraction when genomic DNA is separated along a density gradient (Macaya, Thiery et al. 1976; Thiery, Macaya et al. 1976), due to its distinct buoyant density. Satellites are predominantly localised at the telomeres and centromeres, and there is evidence to suggest that they are nuclease resistant (Yunis, Roldan et al. 1971).

1.1.4 CpG islands

CpG islands are genomic stretches with about ten-fold higher concentration of CpG dinucleotides than the rest of the genome and often exceed 200 base pairs in length. They occur near the sites of transcription initiation of around 50% of all genes, including all housekeeping genes (Bird 1986). Craig and Bickmore reported that T bands had the highest concentration of CpG islands (Craig and Bickmore 1994).

In the human genome it is estimated that 70-80% of CpG dinucleotides are methylated. Unless associated with imprinted or X-linked genes (Riggs and Pfeifer 1992; Razin and Cedar 1994), CpG islands form the exception to the rule and remain unmethylated (Cross and Bird 1995). The colocalisation of these islands with gene promoters is suggestive of a putative regulatory role for CpG methylation in transcription. In early developmental stages a small proportion of islands become methylated, thus contributing to the transcriptional silencing of the associated genes. Nonetheless, the majority of CpG islands remain unmethylated even when the associated gene is transcriptionally inactive (Antequera and Bird 1999), while the mechanism which enables CpG islands to evade methylation still remains unsolved.

Apart from the primary sequence, chromatin architecture is a significant factor contributing to the structural and functional organisation of chromosomes. The interaction of proteins with the DNA determines chromatin conformation. The next four sections discuss issues relevant to this topic.

1.1.5 Nucleosomes and chromatin folding

The 46 chromosomes of a diploid human cell have two metres of DNA, which is efficiently packaged with the help of proteins to fit into the cell nucleus. Despite the immense level of DNA condensation, the various nuclear processes such as transcription, DNA replication and repair, are efficiently accommodated. The complex of DNA and proteins, primarily the histones, as they exist within the cell nucleus, is termed chromatin.

The fundamental units of chromatin are the nucleosomes. Each nucleosome consists of 146 bp of DNA wrapped twice around a di-tetrameric octamer of the four core histone proteins: H2A, H2B, H3 and H4, assuming a "beads on string" configuration that is approximately 11 nanometres in width (Luger, Mader et al. 1997). The histones are arranged such that their amino tails protrude outwards from the nucleosome core, a feature that has immense significance in chromatin remodelling.

Each nucleosome is connected to its neighbours by a short segment of linker DNA. The next level of packaging is the arrangement into a 30 nanometre fibre, also known as chromatin fibre, roughly producing 50-fold compaction (Figure 1.3). Two competing models have been proposed. Finch and Klug initially proposed the solenoid model which has a regular structure with coiling of the linker DNA and formation of super helical turns (Finch and Klug 1976). The alternative model describes an irregular zig-zag pattern folding into a slinky-like structure (Woodcock, Grigoryev et al. 1993; Leuba, Yang et al. 1994). Another protein family, the Linker Histones (LH), provides stability to the 30 nanometre fibre by binding each nucleosome to its adjacent linker DNA. The details of folding that result to levels of structure beyond the chromatin fibre and ultimately form the highest order structures, like the mitotic chromosome, still remain unclear. Nevertheless, various models have been proposed based on observations made by different methodological approaches.

1.1.6 Matrix Attachment Regions

In the early 1960s, while studying lampbrush chromosomes of newt oocytes, Gall and Callan observed DNA-containing asymmetric "loops" arising from the chromosome axis (Gall and Callan 1962). These loops may correspond to the 120 Kb loops described by Münkel and co-workers, which form rosette structures (Munkel, Eils et al. 1999). Each rosette was suggested by Münkel to be attached at its centre to proteins, as shown in work pioneered by Laemmli, that is usually referred to as nuclear scaffold or matrix (Paulson and Laemmli 1977; Marsden and Laemmli 1979). Early studies identified Topoisomerase II (topo II) as a major structural component of the scaffold (Gasser, Laroche et al. 1986). Specific ~200 bp long AT-rich sequences called scaffold or matrix attachment regions, commonly abbreviated to SARs and MARs, anchor the rosettes to the protein scaffold. The concept of the nuclear scaffold or matrix has been controversial as it is considered by some to be a methodological artefact.

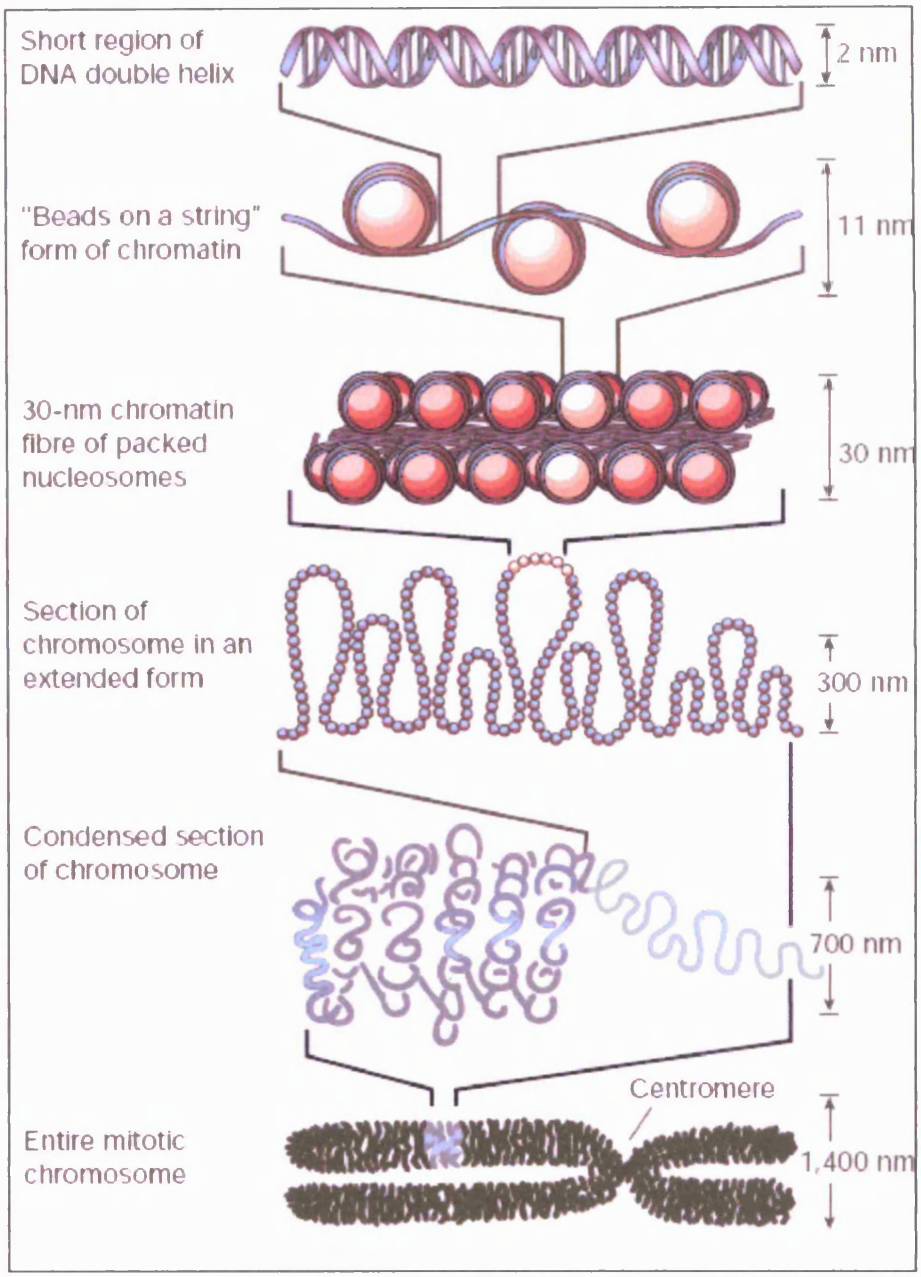


Figure 1.3 Packaging of DNA into chromatin

The highest level of compaction for chromatin is established in mitosis where the chromosomes condense down to 700 nanometres in diameter. This is an ATP dependent process catalysed by a highly conserved pentameric protein, 13S condensin, in the presence of topo I (Kimura, Cuvier et al. 2001). In accordance with previous studies that have linked topo II with the structural organisation of chromosomes, this protein has recently also been implicated in mitotic chromosome condensation (Cuvier and Hirano 2003). As chromosomes condense during mitosis, the SARs across the entire length of the chromosome line up onto the scaffold to form an AT-rich row named AT-queue. It has been proposed that the differential staining of G and R bands is a direct consequence of the compaction of the AT queue. In G bands the AT-queue is made up of tightly folded short DNA loops and is able to generate a strong fluorescent signal. In R bands, on the other hand, the DNA loops are larger and less frequently folded so the AT-queue remains more central and therefore the signal is faint (Saitoh and Laemmli 1994). This variation in folding is also portrayed in the DNase I sensitivity of different genomic regions. R bands usually, though not always correspond to DNase I sensitive segments (Kerem, Goitein et al. 1984).

1.1.7 Different levels of chromatin compaction

Chromatin in interphase cells has different levels of compaction across the genome. This phenomenon had been detected in early electron microscopy studies which identified regions of differentially packaged chromatin within the nucleus. Densely packaged, condensed chromatin is conventionally termed heterochromatin while the more sparsely arranged and decondensed portion is called euchromatin (Felsenfeld and Groudine 2003). Euchromatin is often associated with active segments of chromosomal DNA where the resident genes are expressed. Conversely, heterochromatin predominantly consists of repressed segments with transcriptionally inactive sequences. Specific chromosomal regions such as telomeres and centromeres are constitutively heterochromatic. Euchromatic regions are generally more dynamic and in the process of various cellular events they have the potential to undergo dramatic changes in compaction (Holmquist and Ashley 2006).

1.1.8 Chromatin remodelling proteins

Several proteins have been implicated in chromatin remodelling and compaction, including the histones, the HMGB and HMGN families, and HP1. Some of these proteins can facilitate changes in chromatin compaction as extensive as turning euchromatic regions into heterochromatin and vice versa, or may simply affect the expression of a particular gene or group of genes.

Transcription factors regulate expression by binding to specific regulatory sequences upstream of genes. Although some are accessible by these factors, interaction may not always be possible because most sequences remain concealed within the nucleosome. Chromatin remodelling complexes assist factors to gain access to these sequences by shuffling the positions of nucleosomes and exposing the sequences for short periods of time (Becker and Horz 2002; Narlikar, Fan et al. 2002). Nucleosomes play an active role in this recognition process by accommodating chemical modifications that incite recognition and binding by specific regulatory proteins. When nucleosomes become fixed to their positions they maintain a close conformation and gene expression is repressed.

The histone amino tails protrude outwards from the nucleosome core. Numerous enzymes use the histone tails as substrates for covalent modification in what is known as the histone code (Strahl and Allis 2000). For example, histone acetyl-transferases (HATs) are responsible for attaching an acetyl group to lysine residues (K) at various positions such as histone H3 K9. Except for acetylation, specific histone residues can undergo methylation, phosphorylation and ubiquitilation (Figure 1.4) (Felsenfeld and Groudine 2003). The addition of acetyl groups to the histone tails is associated with opening of the nucleosomes and higher transcriptional activity. Except for methylation on histone H3 K4 that favours gene expression, histone methylation is generally associated with the exact opposite effects of acetylation. It brings about gene silencing and increases chromatin compaction via its suggested involvement in heterochromatin assembly. [See Appendix A for more information on proteins that affect chromatin structure]

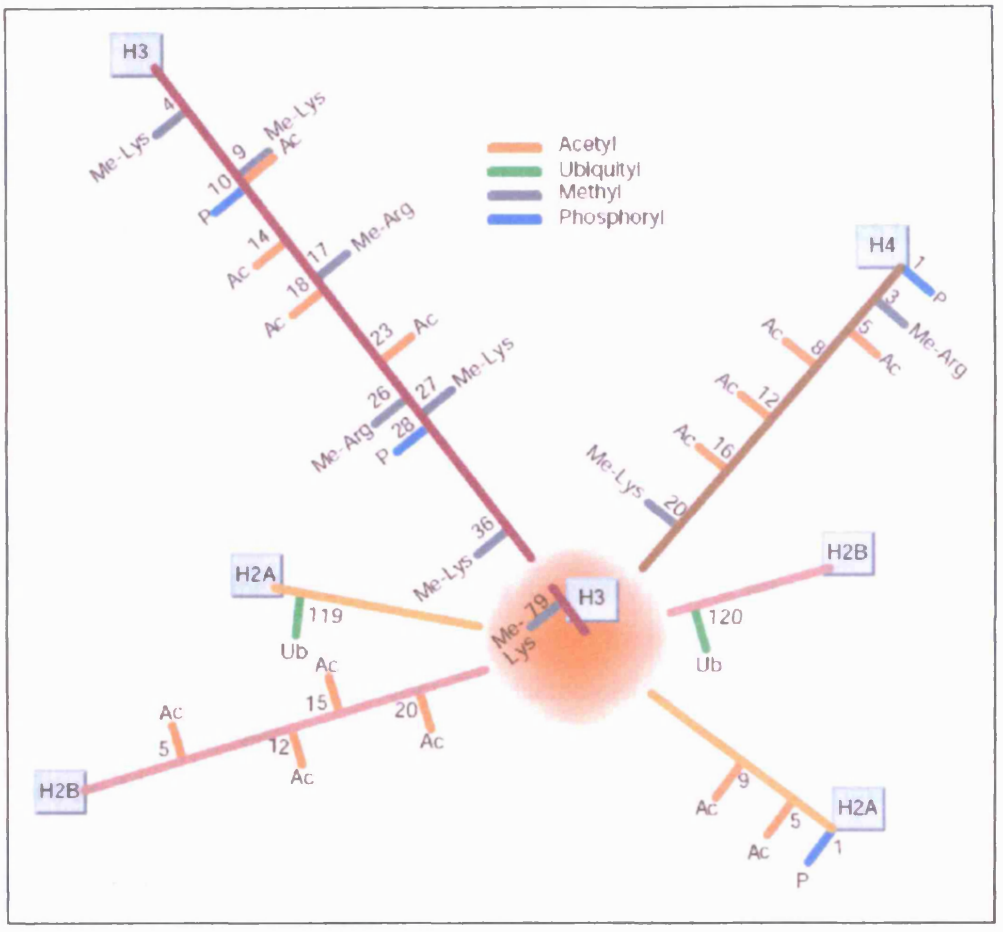


Figure 1.4 Map of histone modifications

Reprinted by permission from Macmillan Publishers Ltd: Nature (Felsenfeld and Groudine 2003) copyright 2003.

1.1.9 Chromosomes and the nucleus

Beyond the characteristics of the primary DNA sequence and its arrangement in chromatin domains as discussed above, the genome has an additional level of organisation, the positioning within the three-dimensional confines of the interphase nucleus. Studies at the beginning of last century suggested that chromatin was not diffused throughout the nucleus (Cremer and Cremer 2006). Since then, various experimental approaches have clearly established that individual chromosomes occupy distinct nuclear domains called territories (Cremer and Cremer 2006).

In interspecies hybrid cell lines containing single human chromosomes in a mouse or hamster background, *in situ* hybridisation with fluorescently-labelled human DNA showed that the human chromosomes occupied discrete domains of the nucleus (Manuelidis 1985; Schardin, Cremer et al. 1985). Chromosome territories (CTs) were subsequently demonstrated within nuclei of human cells, using the approach known as chromosome painting (Lichter, Cremer et al. 1988). In 2005, Bolzer and colleagues established a 3D position map of the preferential positions of all 46 CTs within human diploid nuclei at interphase. Gene-poor chromatin domains were found to layer just below the nuclear envelope, while gene-dense chromatin was found to be enriched in the nuclear interior (Bolzer, Kreth et al. 2005).

Between and branching within these CTs is suggested to be a space which has been named the interchromatin domain (ICD). The ICD forms a network of channels that extend to the nuclear pores and contains macromolecular complexes which have a role in transcription, splicing, DNA replication and repair (Cremer, Kurz et al. 1993; Zirbel, Mathieu et al. 1993). Evidence from confocal microscopy experiments on the active and inactive X chromosome suggested that CTs are not smooth structures (Eils, Dietzel et al. 1996). This suggestion has been supported by the observation of several-Mb long chromatin loops which extended outwards from the surface of the chromosome 6 domain, in response to cytokine treatment (Volpi, Chevret et al. 2000).

1.2 Eukaryotic replication

Every time a eukaryotic cell divides, a perfect copy of the genome must be generated within S-phase. This is achieved via a highly coordinated process that precisely duplicates the genome once during the cell-division cycle. DNA replication initiates at numerous sites - the origins of replication - along the eukaryotic chromosome; it then proceeds in both directions away from those sites until it merges with the DNA that is replicated from the adjacent origin. The length of DNA sequence that falls within the replication domain of a particular origin is termed a "replicon". Some origins become activated before others during the S-phase, consequently replicating their associated domains ahead of others. This temporal distinction is commonly referred to as

replication timing and it has been associated with structural and functional characteristics of chromatin.

1.2.1 Licensing chromatin for replication

Evolution has established control mechanisms that regulate the successful progression through various landmarks, to ensure that the entire genome replicates correctly before the cell cycle can proceed to the next stage. These are the cell cycle checkpoints which are enforced by different regulatory proteins. For example, DNA damage during G1 inhibits entry into S-phase and prevents replication of damaged DNA. Checkpoint activation facilitates repair or induces programmed cell death in case the damage is irreparable.

At the transition of mitosis and G1, DNA becomes licensed to replicate by the introduction of pre-replicative complexes (pre-RCs) at multiple sites across the genome. These complexes consist of protein subunits that have been very highly conserved in the course of metazoan evolution. Assembly of the pre-RC starts by restoring the hexameric origin recognition complex (ORC) at sites near potential origins of replication (Mendez and Stillman 2003). Then Cdc6/18 and Cdt1 (Cdc10 dependent transcript 1) followed by the ring-shaped mini-chromosome maintenance (MCM) complex, consisting of Mcm2-7, are recruited and assembled onto the structure in ATP-dependent steps rendering the origins licensed for replication. Eventually, having fulfilled the requirements of the late G1 checkpoint, the cells enter S-phase. In response to mitogenic growth factors, the G1 type cyclins D and E are induced, activating their catalytic partners, the cyclin dependent kinases (CDKs). Simultaneously, the CDK inhibitors $p21^{Cip}$ and $p27^{Kip1}$ become down-regulated (Pardee 1989; Sherr 1994; Sherr and Roberts 1999). These events lead to the progressive phosphorylation of the retinoblastoma protein (pRb) by the CDKs, so by the end of G1 pRb is hyperphosphorylated and inactive (Weinberg 1995). This marks the passage through the critical cell cycle restriction point, R, after which the cells are committed to progress into S-phase.

1.2.2 Entry into S-phase

Transition to S-phase is concomitant with a dramatic increase in the levels of Cdk2-CyclinE and DDK (Dbf4-dependent kinase), which activate components of the Pre-RCs. This is accompanied by recruitment of the replicative polymerases and other replication proteins that begin DNA synthesis. The formation of new Pre-RCs and re-replication is inhibited by the mitotic CDK (Cdk1) so origin re-licensing can only take place at the next mitosis to G1 transition. Cdt1 disappears as soon as replication is underway and is thought to be regulated by phosphorylation dependent proteolysis by the Skp2-Cullin-F-box (SCF) family of ubiquitin ligases (Nishitani, Lygerou et al. 2004; Sugimoto, Tatsumi et al. 2004). Another mechanism that has evolved in higher eukaryotes to prevent re-replication involves the protein geminin, which binds to and inactivates Cdt1 thus preventing inappropriate licensing (Wohlschlegel, Dwyer et al. 2000; Nishitani, Taraviras et al. 2001; Melixetian, Ballabeni et al. 2004). Mammalian Orc1, a component of the ORC complex, dissociates from chromatin and is degraded in S-phase (Kreitz, Ritzi et al. 2001; Li and DePamphilis 2002). Surprisingly, much of Cdc6/18 remains bound to chromatin throughout S-phase and G2 only to undergo Anaphase Promoting Complex (APC)-dependent proteolysis at the end of mitosis (Coverley, Pelizon et al. 2000; Mendez and Stillman 2000).

1.2.3 S-phase checkpoints

A set of checkpoint proteins are required to deal with endogenous stresses that occur during normal cell proliferation. Three types of S-phase checkpoints exist: (i) the replication checkpoint, which is initiated when the progression of replication forks becomes stalled; (ii) the S-M checkpoint, which ensures that cells do not attempt to divide before the entire genome is duplicated; and (iii) the intra-S-phase checkpoint, which is activated by DNA double strand breaks (DSBs) (Canman 2001; Bartek, Lukas et al. 2004; McGowan and Russell 2004; Li and Zou 2005). The different checkpoints often activate parallel and overlapping pathways that feed into common downstream effectors. As a result, cells defective for checkpoint proteins are highly sensitive to damaging agents and cellular stress. Recent evidence suggests that some

proteins which have traditionally been considered as checkpoint components also play a role in maintenance of replication forks (Shechter and Gautier 2004).

1.2.4 Replication machinery

The precise steps leading to replication initiation are currently poorly understood. Following the licensing of chromatin, a series of events takes place at the origins of replication where the Pre-RCs initiate replication by promoting origin unwinding and facilitating recruitment of the replicative polymerases.

There is evidence to suggest that the DNA polymerases Pol δ (3-4 subunits) and Pol ε (4 subunits) function with the proliferating cell nuclear antigen (PCNA) as monomeric polymerase particles to simultaneously replicate the leading and lagging strands respectively (Figure 1.5). PCNA is a ring-shaped homotrimeric protein and functions as a processivity sliding clamp. PCNA clamps are thought to be left behind on lagging strand fragments as they interact with factors necessary for Okazaki fragment maturation (Kao and Bambara 2003). In eukaryotes the lagging-strand fragments are about 200 nucleotides long. Remarkably the eukaryotic replication fork progresses at a 10-fold slower rate than that of *Escherichia coli*. The replisome consists of several proteins. There is evidence to suggest that the MCM complex corresponds to a helicase function that translocates the DNA on the leading strand (Forsburg 2004). The primase function comes from the smallest subunit of the four-subunit DNA Pol α that synthesises its own RNA primer (~12-nucleotide) and then extends it with about 20 nucleotides of DNA (Lehman and Kaguni 1989). The hetero-trimeric replication protein A (RPA) functions as a single stranded DNA binding protein (SSB), protecting single stranded DNA and melting hairpins (Wold 1997). Other proteins like Cdc45, the GINS complex, Dpb11, Sld2 and Sld3 are also thought to act during replication initiation and fork progression (Kamimura, Masumoto et al. 1998; Kamimura, Tak et al. 2001; Takayama, Kamimura et al. 2003), although their entire spectrum of functions is not yet fully understood. The pentameric replication factor C (RFC) is the eukaryotic clamp loader. It is a DNA-dependent ATPase and is proposed to act as a scaffold bridging Pol δ and Pol ε coupling them to the helicase (Johnson and O'Donnell 2005).

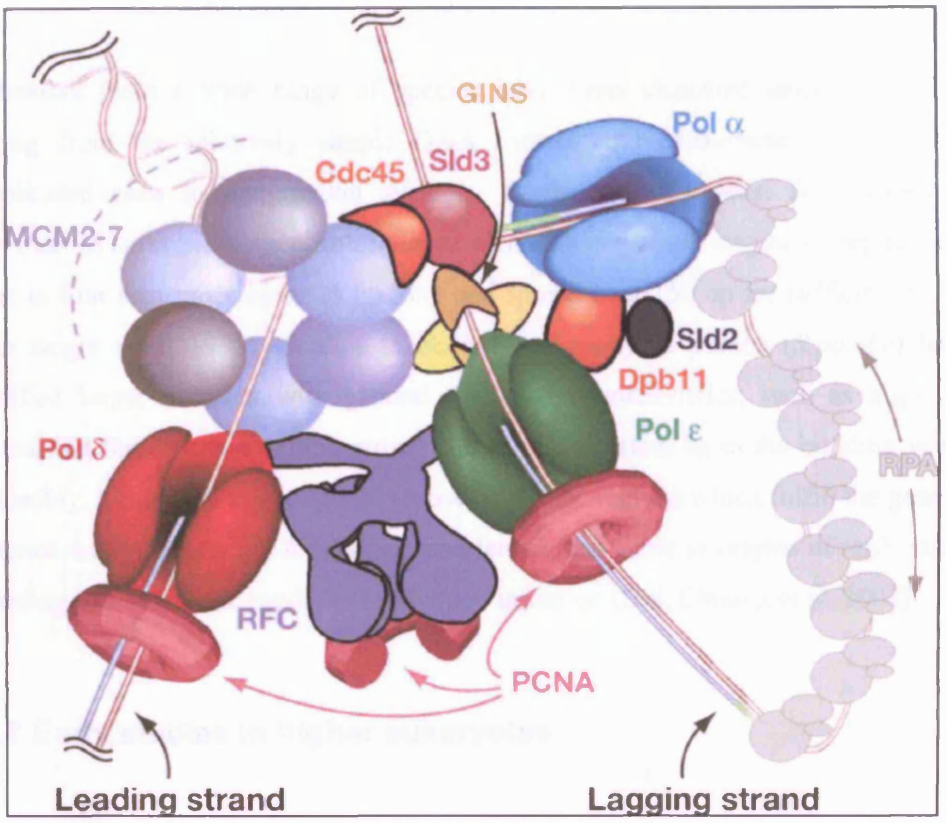


Figure 1.5 Hypothetical arrangement of proteins at the eukaryotic replication fork.

1.2.5 The "Replicon" model

In the early 1960s, Jacob and Brenner proposed a model to explain the regulation of *E.coli* replication (Jacob and Brenner 1963). This widely known Replicon model stated that the replicator, a genetic element determining replication initiation sites on the DNA, was the target of a regulatory factor, termed "initiator". The interaction of the replicator and the initiator could activate initiation to replicate a circular structure. This simple idea was a significant landmark in the field of DNA replication, and the principle was developed to describe the replication of eukaryotes.

1.2.6 Replication origins in yeast

Replicators from a wide range of species have been identified since the 1960s, ranging from the relatively simple DNA viruses and prokaryotes to the more complicated ones in mammalian systems. In the budding yeast *Saccharomyces cerevisiae* (*S.cerevisiae*) specific sequence elements constitute origins of replication. Three to four sequences up to 15 bp long and spread over 150 bp are sufficient to act as an origin (Bell 1995). Studies in *Schizosaccharomyces pombe* (*S.pombe*) have identified larger domains with general sequence characteristics, such as high AT content, but these do not exhibit strong sequence consensus as in the budding yeast. Noticeably, it has been suggested that most intergenic regions which fulfill the general sequence requirements, i.e. AT content and length, can serve as origins of replication, according to a stochastic model of replication initiation (Dai, Chuang et al. 2005).

1.2.7 Early studies in higher eukaryotes

In higher eukaryotes and more specifically in human and other mammals, the situation is more complicated than in yeast. Not only are the replicators sequence unspecific, but also at least two different mechanisms of operation apparently exist: the origins of replication, and the regions or zones of initiation.

Soon after the Replicon model was proposed, Huberman and Riggs used fibre autoradiography to show that replication initiates at multiple sites on mammalian chromosomes and that the replicators have an average spacing of 100 Kb (Huberman and Riggs 1968), while the replication fork was estimated to proceed at an average rate of 0.3-6 Kb per minute (Edenberg and Huberman 1975). In some cases, however, similar studies showed that replicons can reach up to an estimated length of 1 Mb (Yurov and Liapunova 1977). The observation that the replicon size can vary was also confirmed for the six replicons of the human dystrophin gene, the first gene to be identified that is replicated by more than one replicon. The investigation was carried out with a replication direction assay which showed that some of the replicons are highly asymmetrical. Remarkably, a termination site, where two replicons merge, was mapped near the location of a recombination hotspot (Verbovaia and Razin 1997).

1.2.8 The ORC complex at sites of replication initiation

The human ORC complex is an essential factor for replication initiation in human cells and it is known to bind preferentially to AT-rich DNA. Interestingly, *in vitro* experiments have shown that it does not discriminate between sequences from characterised human origins and other control sequences; it was also able to stimulate initiation from any DNA sequence (Vashee, Cvetic et al. 2003; Schaarschmidt, Baltin et al. 2004). These observations have provided an explanation for the apparent promiscuity of replication initiation in human and other mammals. Additionally, the affinity of ORC for negatively supercoiled DNA is considerably greater than its affinity for linear or relaxed DNA, according to a study performed in *Drosophila melanogaster* (Remus, Beall et al. 2004).

1.2.9 Studies on mammalian replicators

In recent years, several mammalian replicators have been identified either as specific origins of replication or as initiation zones. The fundamental difference between the two is that an origin specifies an initiation event within 0.5-2 Kb, while an initiation zone within 10-50 Kb, contains several sites with the potential to function as origins of replication. Only one of the sites within a zone is most likely to be triggered during the S-phase in each cell. Furthermore, one of those sites is used more frequently than the rest.

In the Chinese hamster dihydrofolate reductase (*DHFR*) locus there is a broad zone of potential initiation sites dispersed throughout an intergenic region spanning 55 Kb. It contains a minimum of 20 potential origins of different activation efficiency (Wang, Dijkwel et al. 1998; Dijkwel, Wang et al. 2002). Three specific sites within this region *oriβ*, *oriβ'* and *oriγ* are preferentially used. A strategy to create various deletions was taken in an attempt to identify whether any particular region within the 55 Kb was essential for replication. Chemical synchronisation of the cells followed by analysis of replication intermediates on 2-D gels, however, showed that none of the deletions

affected the efficiency of replication at the *DHFR* locus. This suggests that the deleted segments do not contain essential elements for replication initiation (Mesner, Li et al. 2003).

Replication at the human β-globin (*HBB*) locus is bi-directional, it is initiated at a single origin located upstream of the *HBB* gene itself, and is used in both expressing and non-expressing cells. Deletion of the origin sequence abrogates initiation from that site and leads to passive replication of the region from further upstream (Kitsberg, Selig et al. 1993). The locus control region (LCR) is 50 Kb upstream of the *HBB* origin and regulates the transcription and chromatin structure of the entire locus. In a deletion mutant lacking the LCR, replication initiated in the 3′ to 5′ direction, suggesting that replication initiation in this locus may require interaction with distal elements like the LCR (Aladjem, Groudine et al. 1995).

The most extensively studied origin of replication is that located immediately downstream of the lamin B2 (*LB2*) locus. The genomic segment encompassing the origin was initially identified from a library of nascent DNA sequences obtained from synchronised cells as they entered S-phase (Tribioli, Biamonti et al. 1987). Subsequently, initiation activity was localised to a ~500 bp region spanning the 3′ end of the *LB2* gene using a PCR method (Giacca, Zentilin et al. 1994). *In vivo* footprinting analysis showed that there is strong protection over a sequence that coincides with the locus where initiation activity had previously been identified (Abdurashidova, Deganuto et al. 2000). The protection is always present for a "core" sequence but extends further out depending on the stage of the cell cycle. When a 1.2 Kb fragment of the *LB2* origin was integrated ectopically in the genome it functioned as a replicator. Notably, a nearby CpG island that was partly included in the fragment was able to influence the replicator (Paixao, Colaluca et al. 2004).

1.2.10 Replication in early stages of development

The fertilised *Xenopus* egg undergoes twelve synchronous rounds of cell division in only 7 hours, a much shorter period than the 36-hour somatic cell cycle. Early studies analysing the replication of ribosomal DNA (rDNA) in *Xenopus leavis* embryos

before the mid-blastula transition, found that the average replicon is around 10 Kb in length and replicates within about 7 minutes. Intriguingly, this period was similar to the duration of the entire S-phase, while the length of DNA was the maximum that could be replicated from a single origin within that time. These observations led to the inference that most rDNA origins must be synchronously activated at the onset of S-phase and that they must be evenly spaced such that no segment remains unreplicated in the end (Hyrien and Mechali 1993). The same authors later reported that when transcription at the rDNA genes resumed in the late blastula/ early gastrula stage of development, replication initiation within the expressed genes was repressed. Nonetheless, the frequency of initiation in the intergenic regions remained constant suggesting perhaps that in metazoan development origin usage becomes confined due to chromatin remodelling and transcriptional activation (Hyrien, Maric et al. 1995). An investigation into the replication of *Xenopus* sperm DNA in *Xenopus* egg extracts found that most origins were spaced 5-15 Kb apart and that they were organised in small, synchronously-firing clusters. When ORC availability was limited, the origin spacing increased suggesting that ORC binding determines the regular spacing of origins in that system (Blow, Gillespie et al. 2001).

1.2.11 The Origin Decision Point

Observations made in the studies described above and elsewhere have established that chromosome replication in *Xenopus* and *Drosophila* embryos is very rapid and utilises non-specific initiation sequences. It is vital, however, for the spacing of these numerous initiation sites to be at regular intervals in order to replicate the entire genome on time. In response to protein kinase signalling pathways, the cell nucleus is required to make a choice for origins (Wu and Gilbert 1996). A discrete transition point, the origin decision point (ODP), has been identified early in G1, which is believed to be the stage at which origins are chosen and replicon size is determined.

Two competing models have been proposed to describe the spacing of origin licensing and timely replication of the genome in the embryos of at least these two metazoans, in what is known as the "random completion problem". According to the first model, the ODP-chosen origins are spaced at regular intervals and each one fires with very

high efficiency. The second model states that a large excess of origins is licensed at the ODP and at the ensuing S-phase just a fraction of those are initially activated, to be followed by firing of origins in unreplicated gaps. Recent evidence appears to offer support to the second model (Hyrien, Marheineke et al. 2003). The assumption that there are more licensed sites on the chromatin than those used as origins raises a potential problem. The unused pre-RCs will eventually become replicated passively by progressing forks from other origins and therefore have to be inactivated in order to avoid re-replication of those sites. Components of the replication checkpoints are likely involved in origin inactivation (Santocanale and Diffley 1998; Santocanale, Sharma et al. 1999).

1.2.12 Factors defining origin selection

Work on *Xenopus* and *Drosophila* embryos offers a paradigm for the frequency of pre-RC formation and origin selection in early stages of metazoan development. Various parameters mediate these early stages of apparently random choice to determine site specificity on the chromatin. The nucleus, for example, has a role in initiation site selection because it concentrates the necessary replication factors within a fixed space.

The structure of chromatin has a significant role in origin activity. Changes in chromatin structure affect the accessibility of initiation factors and can create new initiation sites or eliminate others. For instance, histone H1, which is generally associated with closed chromatin structure, has been shown to cause a reduction in the frequency of initiation (Lu, Sittman et al. 1998). Further evidence for the importance of chromatin structure in origin activity was provided by experiments with TSA. Treatment of HeLa cells with this histone deacetylase inhibitor promoted a more dispersive pattern of initiations, while the late firing origin at the *HBB* locus initiated replication earlier under those conditions (Kemp, Ghosh et al. 2005). These findings suggest that the variable activity of replication origins observed throughout the genome is related to the structure of chromatin.

Several studies suggest that DNA sequence is important in origin specification and function, including those on the *HBB* origin mentioned above. Deletion of the origin sequence in the *HBB* locus inhibited replication initiation, whereas deletion of the LCR distorted the origin function. Notably, the same origin is used regardless of *HBB* gene expression (Kitsberg, Selig et al. 1993; Aladjem, Groudine et al. 1995), indicating that origin activity is irrelevant of expression status at this locus. Similarly, the same origins of replication are used on both the active and inactive X-chromosome (Cohen, Brylawski et al. 2003; Gomez and Brockdorff 2004). An investigation into the ability of specific replicator sequences to initiate replication at ectopic chromosomal regions, outside their normal environment, has shown that the hamster *DHFR ori-β* retains its endogenous function unless specific sequences are mutated (Altman and Fanning 2001). Reciprocal studies using the human *c-myc* and *LB2* replicators produced analogous findings (Liu, Malott et al. 2003; Paixao, Colaluca et al. 2004). Another illustration of sequence importance includes the observation made by footprinting studies at the *LB2* origin where a specific sequence of DNA was protected in a cell-cycle dependent manner, presumably due to ORC binding (Abdurashidova, Deganuto et al. 2000). Overall, these data demonstrate that metazoan origins of replication are determined by specific sequences near the site of initiation while origin activity can be regulated from sequences such as locus control regions that can be further away.

Active genes generally replicate early and inactive ones replicate late in S-phase. This is suggestive of a role for transcription in origin function. A large proportion of metazoan origins are situated near the promoter region of a gene. The rDNA paradigm mentioned above is an indication of the inhibitory role that transcriptional activation can have by preventing replication initiation within active genes.

A substantial body of evidence suggests that CpG islands play a significant role in origin selection. At least three Chinese hamster genes: *APRT*, *GADD* and *TK*, with characterised active origins of replication within their promoter, encompass a CpG island (Delgado, Gomez et al. 1998). In addition, CpG islands have been found within or proximal to replication origins at the promoters of the human genes: *c-Myc* (Vassilev and Johnson 1990), *hsp70* (Taira, Iguchi-Ariga et al. 1994) and *LB2* (Paixao, Colaluca et al. 2004). Parallel chromatin immuno-precipitation (ChIP)

studies against components of the human ORC complex have identified significant binding to DNA fragments with properties typical of CpG islands (Keller, Ladenburger et al. 2002; Ladenburger, Keller et al. 2002). A mechanism has been proposed to explain how CpG islands evade methylation by the DNA methyl-transferase (MTase), which is diffusely-localised in non S-phase cells, but targets sites of DNA replication during S-phase. Methylation evasion is possible if CpG islands are replicated very early in S-phase, before the MTase becomes associated with the replication machinery (discussed by Delgado, Gomez et al. 1998)). In some other cases, however, regions with methylated CpG dinucleotides have been found in close proximity to sites of initiation, as in the hamster *DHFR ori-β*, which is less likely to fire when the adjacent CpG methylation is removed (Rein, Kobayashi et al. 1999). Taken together, these data do not clarify the role that DNA methylation and CpG islands play in replication initiation. Future studies could offer us more insights into this subject.

Lastly, a recent report has drawn attention to the availability of nucleotides in the process of origin decision and use. Anglana and co-workers focused their studies on the *oriGNAI3* locus in Chinese hamster, which contains a previously mapped site of replication initiation (Toledo, Baron et al. 1998; Svetlova, Razin et al. 2001). They used a DNA combing method to confirm the previous findings and also to examine whether other sites of initiation were present in that region. They identified five secondary initiation sites which are generally suppressed by the primary site and according to their findings these remained inactive when the nucleotide pool available was increased by adding DNA precursors to the culture medium. Conversely, when the nucleotide pool was restricted by addition of hydroxyurea, a specific inhibitor of ribonucleotide reductase, the rate of fork progression decreased as did the efficiency of initiation from the primary *oriGNAI3* locus, while the dormant secondary origins were activated and the frequency of initiation at those sites increased (Anglana, Apiou et al. 2003). The authors suggest that nucleotide availability can have a significant effect on the determination of replication initiation sites and their frequency of distribution within chromosomes. This development could help to bridge previous inconsistencies concerning the presence or not of initiation zones, such as the one described by Vaughn *et al* (Vaughn, Dijkwel et al. 1990) for the hamster *DHFR* locus in contradiction to two other reports (Handeli, Klar et al. 1989; Burhans, Vassilev et

al. 1990). The findings of Vaughn and others that involve chemical synchronisation with hydroxyurea as part of the methodological approach might in fact be a by-product of nucleotide restriction.

In various review articles and other publications Melvin DePamphilis has suggested that the initiation of DNA replication in metazoan chromosomes follows the Jesuit dictum that "many are called, but few are chosen" (DePamphilis 1993; DePamphilis 1998; DePamphilis 1999; DePamphilis 2003). From the widespread and apparently random initiation sites that have been seen in early developmental stages various levels of selection intervene in the adult cell to silence some origins and authorise others. The final product of site-specific initiation is the result of several parameters including nuclear localisation, gene transcription, DNA methylation and local chromatin structure, the proportion of initiation proteins to DNA, the levels of nucleotide pools, and perhaps others that are yet undefined.

1.2.13 Replication timing

When a mammalian cell commits itself to DNA replication, it enters S-phase, where several origins of replication become active and start replicating their respective replicons. However, not all the origins fire at the same time. It is believed that different groups or clusters of origins fire simultaneously, in a process that is ongoing throughout S-phase. As a result, different sections of the genome become replicated before others, and are thus assigned an earlier timing of replication. The replication timing of a particular region is generally fixed and can be reproduced over several cell cycles. Nevertheless, under certain circumstances the timing of replication has been shown to change.

1.2.14 Replication banding and replication across cytogenetic bands

Early studies investigating the replication timing of mammalian chromosomes employed a method of cell synchronisation in G1 followed by labelling during the S-

phase with bromodeoxyuridine (BrdU). This is an analogue nucleotide to thymidine which is incorporated into newly synthesised DNA during replication. It can then be visualised under a fluorescence microscope with the use of a fluorochrome-conjugated antibody that recognises a specific epitope on this molecule. Addition of BrdU to the culture medium in which the cells are grown, at different time-points after release from block, produces different banding patterns on metaphase chromosomes. The genomic regions which replicate early in S-phase become labelled when the cells are pulsed with BrdU soon after release from block, and those which replicate late became labelled when BrdU is administered at a later stage in S-phase (Holmquist, Gray et al. 1982).

A study of Chinese hamster fibroblasts using BrdU incorporation revealed that chromosomal R bands replicate in the first half of S-phase while the G bands replicate in the second half (Holmquist, Gray et al. 1982). Another study that adopted a similar approach to examine only the early replicating segments of human chromosomes by labelling with BrdU for the first 20 minutes within S-phase found that early replicating regions generally mapped within R bands although not all the R bands had incorporated BrdU within that time frame. It was found possible to define how early a band replicates within the S-phase by the frequency of labelling with BrdU. The most frequently labelled band was 15q22, leading the authors to suggest that it contains some of the earliest replicating sequences in human fibroblasts (Cohen, Cobb et al. 1998).

Strehl and co-workers used a PCR based-assay across the early replicating band 13q14.3 and the late replicating band 13q21.1 in an attempt to dissect replication timing over a 4-5 Mb segment that has been proposed to overlap with an R/G band boundary. They found that replication timing across this region shifted gradually from early to late over many Mb, without discrete transitions between them (Strehl, LaSalle et al. 1997). Another PCR-based study using a male cell line examined a 3 Mb region that included the transition between bands Xq13.1 and Xq13.2, an R band and a G band respectively, and neighbouring sequences. This region coincides with the human X-inactivation centre (XIC) and includes the XIST locus. The replication timing was found to undergo a transition from earlier to later in two occasions across the region studied, giving rise to a 1 Mb late replicating zone. The authors suggest that this

corresponds to G band Xq13.2 (Watanabe, Tenzen et al. 2000). Using a fluorescence *in situ* hybridisation (FISH) based assay, in another study they examined the cytogenetic band boundary regions surrounding Xp22.2 which has been described as a darkly staining G band. The results indicated that the replication timing domains were generally consistent with the cytogenetic banding patterns, such that the Xp22.2 band was later replicating than the neighbouring Xp22.1 and Xp22.3 bands. Nonetheless, a sharp replication timing transition was not observed across any of the cytogenetic band boundaries in this case either (Bilyeu and Chinault 1998).

The movement of the replication fork reportedly becomes slower around early to mid S-phase, as it proceeds at the R/G band boundary, according to experiments performed using pulse-chase-pulse labelling on different mammalian cultured cells (Takebayashi, Sugimura et al. 2005). This concept was investigated further with FISH at the boundary between the R and G bands: 1q31.1 and 1q32.1. Takebayashi *et al.* observed a switch in replication timing across this region, which is known to contain a marked decrease in GC content. To understand the relationship between replication timing and GC content, Eyre-Walker analysed sequence and replication data from around 40 genes of various gene-containing regions of the human and mouse genomes, belonging to different isochores. Genes with both high and low GC content were found to replicate either early or late in S-phase (Eyre-Walker 1992).

Using a PCR based assay Brylawski and co-workers determined the replication timing of a 340 Kb region in 1p36.13, an R band. This segment contained a 140 Kb section that replicates within the first hour of S-phase, while the flanking sequences replicated 1-2 hours later. They were also able to characterise two origins of replication that were activated with 1 hour difference from each other. One of the two fires early in S-phase and belongs to the 140 Kb segment while the other one belongs to the neighbouring region that replicates slightly later. Taken together these data suggest that the R band 1p36.13 consists of regions that replicate at different times in S-phase (Brylawski, Cohen et al. 2004).

Thus overall, there is a general a correlation between cytogenetic bands and replication timing. R bands generally replicate early in S-phase and G bands later, while the boundary between these is apparently accompanied by a replication timing

transition that can vary in length and definition. However, even within the same band different sections can replicate at different times.

1.2.15 Asynchronous replication

Further clues to other possible determinants of replication timing came from studies of asynchronously replicating regions, such as imprinted genes, the inactive X chromosome in mammalian females and other monoallelically expressed sequences. Imprinting is a phenomenon associated with some genes of our genome where only the allele from the paternally or the maternally-inherited chromosome becomes expressed in the offspring, while expression from the other allele is silenced.

To examine the effect of imprinting on chromosomal replication, Drouin and co-workers analysed bands having one or more imprinted genes by BrdU banding (Drouin, Boutouil et al. 1997). Their assay was similar to that of Holmquist, described above. The imprinted genes did not induce delayed replication in the bands where they reside. This finding suggests that potential changes in the replication timing of single genes might not be visible at this resolution.

To analyse replication timing in imprinted loci, several studies using fluorescence *in situ* hybridisation have been performed. Kitsberg *et al* investigated the replication timing of mouse and human imprinted genes including: *Igf2*, *Igf2r*, *H19* and *Snrpn*. The actively expressed paternal alleles were always found to replicate earlier of the two homologues (Kitsberg, Selig et al. 1993). Some of these findings were confirmed in a study examining the replication timing of sections of human chromosome 11, where the imprinted genes *Igf2* and *Wt1* reside. The paternal *Igf2* replicates early, while *Wt1* has an early replicating maternal allele (Bickmore and Carothers 1995). Another study focusing on the mouse chromosome 7 distal imprinting cluster found that a 1 Mb region between the genes $p57^{Kip2}$ and *H19* replicates asynchronously, the paternal allele replicating earlier (Kagotani, Takebayashi et al. 2002).

Initial evidence suggested that asynchronous replication of imprinted loci is due to DNA methylation at or near the promoters of the inactive alleles (Bickmore and

Carothers 1995). Additionally, a role has been proposed for histone acetylation, overall suggesting that imprinting-associated replication reflects differences in chromatin structure between the two homologous alleles (Bickmore and Carothers 1995; Kagotani, Takebayashi et al. 2002). Replication asynchrony is not detected in tissues with low expression at the active alleles of imprinted loci (Kawame, Gartler et al. 1995). Replication timing for the *SNRPN* gene had been examined in highly expressing lymphocytes, where one allele is very highly expressed, and was found to be asynchronous, whereas in HeLa cells where expression is low, asynchrony was less pronounced. This suggests a role for the level of expression from the active allele in the timing of replication.

Another group of asynchronously replicating sequences are the randomly monoallelically expressed genes. For example, in olfactory sensory neurones, one gene from a family of over a thousand odorant receptor genes is selected and only one of the two alleles actually becomes expressed (Chess, Simon et al. 1994). The two alleles replicate asynchronously. A similar pattern of replication behaviour is detected in the three B-cell-receptor loci: μ, κ and λ, and the T-cell-receptor β (*TCRβ*) locus. Asynchronous replication in these genes has been associated with the phenomenon of allelic exclusion, such that one of the two alleles becomes expressed after being randomly established at a stage in early development (Mostoslavsky, Singh et al. 2001). The phenomenon where one of the two alleles is randomly expressed has been studied for various such loci that exist on the same pair of homologous chromosomes. In mice, it had initially been found that the alleles of all tested genes with asynchronous replication reside on the same chromosome, randomly selected in different cells to be either the maternal or paternal copy (Singh, Ebrahimi et al. 2003). Similar findings were obtained in a subsequent study with human cells. It has been suggested that a mechanistic link exists between this "chromosome pair non-equivalence" and dosage compensation (Ensminger and Chess 2004).

1.2.16 Replication timing at the X chromosome

Chromosome pair non-equivalence is a feature associated with the X-chromosome in mammalian females, where X inactivation occurs in early embryos. A large

proportion of the inactive X becomes condensed and the majority of the residing genes are not expressed. The silenced state is clonally inherited through subsequent cell divisions. Almost the entire inactive X-chromosome replicates later than the active one, although some discrepancies have been reported. There are conflicting reports for the replication timing of the *XIST* gene, which is expressed on the inactive X and silenced on the active X-chromosome. Some studies have provided evidence that the expressed *XIST* allele on the inactive X replicates earlier than the silent allele (Hansen, Canfield et al. 1995; Xiong, Tsark et al. 1998), while others have suggested quite the opposite (Torchia, Call et al. 1994; Torchia and Migeon 1995; Gartler, Goldstein et al. 1999).

The pseudoautosomal region PAR-1 at the terminal 2.6 Mb on the short arm of chromosome X shares sequence homology with the equivalent section of the Y chromosome. Its constituent genes escape X-inactivation and are expressed on both X chromosomes in females. The boundary of the pseudoautosomal region, which has been assigned to Xp22.33, has provided an interesting model to study replication timing because of the potential sharp transition in chromatin arrangement on either side. In a male derived cell line the entire PAR-1 region behaves as a domain of uniform replication (Bilyeu and Chinault 1998). Sharp transitions in replication timing are not present at the boundary of this region. These findings suggest that the cytogenetic band boundaries may represent a transition in some aspect of the DNA or chromatin that appears to be sharp at the cytogenetic level but is gradual at the molecular level.

Another study focusing on the replication timing of the inactive X-chromosome suggests that this replicates late as a consequence of the gene silencing and heterochromatin formation (Gomez and Brockdorff 2004). Interestingly, in the same study, origin usage is not affected by the closed chromatin structure in the inactive chromosome, since both the active and inactive X have been found to fire all the same origins that had been investigated.

1.2.17 A link between replication and transcription

The findings described above suggest a relationship between transcription status and replication timing. This finding is supported by studies using array technology to gather information for large genomic regions, such as chromosomes and entire genomes. The analysis of replication timing in the *Drosophila* genome showed a strong correlation between replication in early S-phase and transcriptional activity (Schubeler, Scalzo et al. 2002). A second study on *Drosophila* identified a temporal pattern that correlates with the density of active transcription. It also found that the effect of transcription status on replication timing is applied over large domains rather than at the level of individual genes (MacAlpine, Rodriguez et al. 2004).

An investigation focusing on chromosomes 21 and 22 in a human cancer cell line found that early replication corresponds to higher gene expression (Jeon, Bekiranov et al. 2005). A separate examination into the replication timing of the human chromosome 22 with two different cell types found that, in general, expressed sequences replicate earlier than silent sequences (White, Emanuelsson et al. 2004). The differential expression of the *IGLL3* locus, which is expressed at much higher levels in lymphoblasts than in fibroblasts, provided evidence in support of the above finding, since it replicates much earlier in the lymphoblastoid cells. Surprisingly, however, White *et al.* identified some regions with annotated expressed genes and others with unannotated transcription units that replicate late. Replication timing analysis of the entire human genome, using a 1 Mb array, found that early replication correlates with the probability of gene expression. Therefore, regions with a high proportion of expressed residing genes are more likely to be early replicating (Woodfine, Fiegler et al. 2004).

Other studies on specific genomic loci have reported similar findings to those generated by array analyses. According to experiments with DNA injection into S-phase nuclei, the DNA is a better template for transcription when the nuclei are in early rather than late S-phase (Zhang, Xu et al. 2002). It has been suggested that the DNA injected into late S-phase nuclei is packaged into chromatin containing deacetylated histone H4, which does not promote an open chromatin state. An

investigation on mouse embryonic stem cells found that DNA replication at some of the neural-associated genes examined becomes early upon neural induction and activation of those genes. In contrast, genes associated with pluripotency replicate early, switching to a later replication program after differentiation. The acetylation level of histone H3, at residue K9, and histone H4 is substantially reduced at the gene *Rex1* (Perry, Sauer et al. 2004). During B-cell development, origin clusters spanning hundreds of Kb become activated within the murine *Igh* locus, coincident with changes in chromatin structure and transcriptional activity (Norio, Kosiyatrakul et al. 2005).

The relationship between transcription of a transgene and the replication timing of its genomic environment was analysed using a human *HBB* transgene whose expression was determined by the orientation of its insertion in the mouse genome (Lin, Fu et al. 2003). The late replicating target site assumed an early timing of replication when the inserted transgene was in the transcription-permissive orientation, and remained late when orientation resulted in silencing of expression. The acetylation levels of histone H3 and histone H4 were also found to change from high to low between the early and late replicating transgenes (Lin, Fu et al. 2003). Findings from previous studies that early replication in the human *HBB* gene is associated with nuclease sensitivity within the region (Dhar, Mager et al. 1988; Cimbora, Schubeler et al. 2000) suggested that open chromatin is associated with early replication. Another investigation with sequences from the human *HBB* locus both in their endogenous context and in transgenic mice showed that elements within the locus control region define early and late replication independently of gene transcription (Simon, Tenzen et al. 2001).

Overall, most studies offer evidence for a relationship between early replication and gene expression or late replication and gene inactivity. The link between replication timing and transcription often relates to chromatin structure and compaction. Early replicating regions have an open chromatin, transcription-permissive structure, whereas compact chromatin is characteristic of late replicating, transcriptionally silent regions. As outlined above, these relationships are not always adhered to.

1.2.18 Replication within the nuclear compartment

Chromosomal DNA replication is temporally and spatially ordered in the nuclei of mammalian cells. As many as one hundred replicons are coordinately replicated at distinct nuclear positions that have been visualised as foci within the S-phase nucleus. It has been suggested that replicons within each replication focus may vary in size and number. DNA replication in each of these foci is believed to be completed in less than one hour (Jackson and Pombo 1998; Ma, Samarabandu et al. 1998).

Distinctive spatial distribution patterns of replication foci have been identified in mammalian nuclei at different stages of S-phase (Nakamura, Morita et al. 1986; Nakayasu and Berezney 1989; van Dierendonck, Keyzer et al. 1989; O'Keefe, Henderson et al. 1992). Replication starts at numerous small foci that are distributed evenly throughout the nucleus. This diffuse pattern corresponds to gene-rich and transcriptionally active regions, analogous to the R/T bands on metaphase chromosomes (Sadoni, Langer et al. 1999). As S-phase progresses the foci become larger and localise near the nucleolus and the nuclear periphery. In the latter stages of S-phase, replication is limited to a few very large foci deep within the nuclear interior (Manders, Stap et al. 1992). These regions correlate with gene poor, heterochromatic regions (Sadoni, Langer et al. 1999).

Replication foci appear to remain at stable positions after they have "fired", with new ones forming *de novo* as S-phase progresses (Leonhardt, Rahn et al. 2000). Evidence that they are associated with an underlying nuclear structure first came from experiments by Hozak *et al.*, who used electron microscopy to visualise discrete replication machinery attached to the nuclear matrix (Hozak, Hassan et al. 1993). Further support for this hypothesis came from a study in which, following matrix preparation, cells were able to synthesise DNA at replication sites. The distribution pattern of replication foci resembled that of intact cells (Nakayasu and Berezney 1989). In addition to all the necessary factors for DNA replication, the immobile foci have been shown to contain proteins involved in cell cycle control (Cardoso, Leonhardt et al. 1993).

1.3 The Major Histocompatibility Complex

The Major Histocompatibility Complex (MHC) is a highly gene rich genomic region on the short arm of human chromosome 6 (Figure 1.6). It is also known as Human Leukocyte Antigen (HLA) complex due to the discovery that certain gene products from this region are localised on the surface of leukocytes (white blood cells) (Dausset 1958; Payne and Rolfs 1958; Van Rood, Eernisse et al. 1958; Payne, Tripp et al. 1964). These genes have been assigned a name prefixed with HLA. The HLA genes are the most polymorphic ever to be discovered, with nearly 2300 alleles identified to date (http://www.ebi.ac.uk/imgt/hla/stats.html).

The biology of the MHC and its gene products have been studied in several species over the years since its discovery in the 1930s, but more extensively in human and mouse. In early studies, the gene products of this region were found to elicit an immune response after allogeneic transplantation (Rapaport and Dausset 1968; Batchelor and Joysey 1969; Ting and Terasaki 1974). In the absence of such reaction there is histocompatibility with the host. Importantly, the MHC contains genes whose products play a vital role in triggering an immune response against infectious agents (Horton, Wilming et al. 2004; Trowsdale and Parham 2004). In addition, this is the most significant region with respect to autoimmune diseases (Larsen and Alper 2004). Intriguingly, it has also been suggested that in some species the MHC gene composition influences mate selection (Jordan and Bruford 1998; Milinski, Griffiths et al. 2005).

The MHC has conventionally been subdivided into three classes according to the biological functions clustered in each one (Horton, Wilming et al. 2004). The MHC classes I, III and II appear in this order from telomere to centromere along the chromosome (Figure 1.6). It is of particular interest that the clustered genes exhibit a tissue-specific expression pattern and are inducible in non-expressing tissues by cytokines such as interferon gamma (IFN-γ). Briefly, the MHC classical class I molecules present endogenous antigens to $CD8^+$ cytotoxic T cells, while class II molecules present exogenous antigens to $CD4^+$ T helper cells. In addition to the

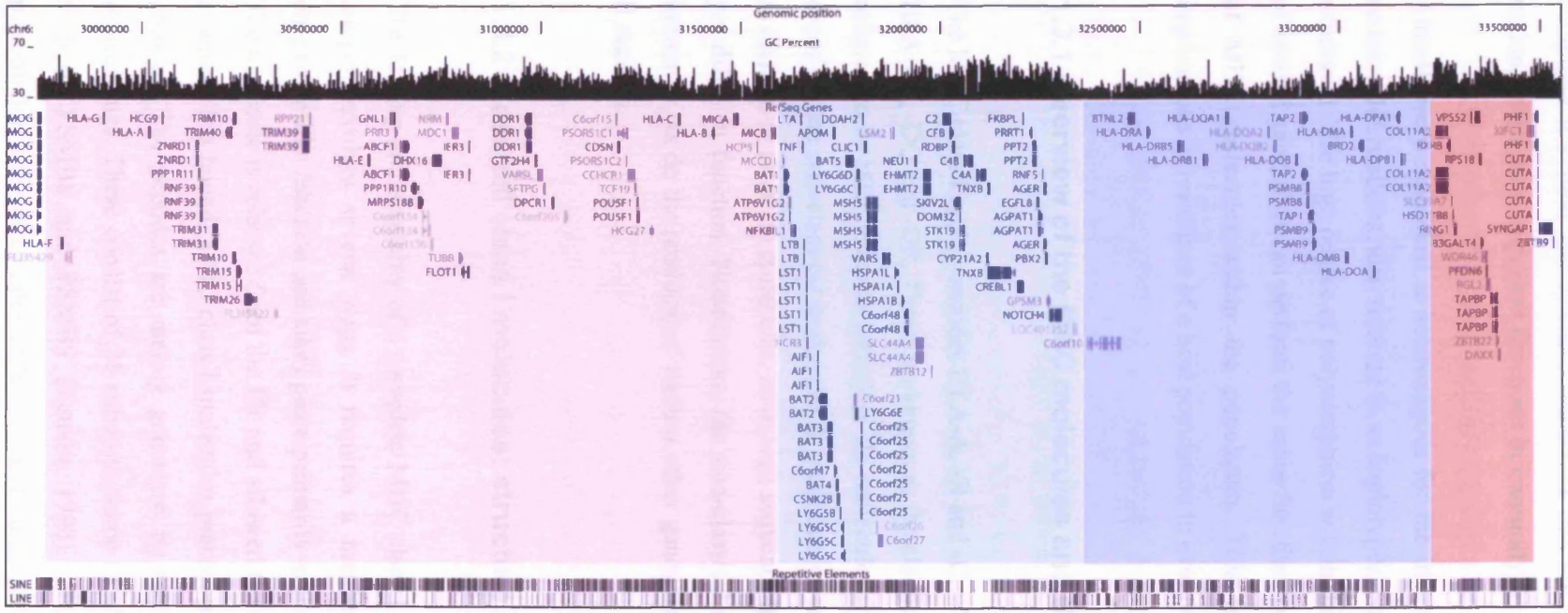


Figure 1.6 Gene map of the human MHC.

The human MHC spans ~4 Mb on the short arm of chromosome 6. Gene annotation according to the RefSeq database, SINE and LINE repeat annotation based on RepeatMasker database (www.genome.ucsc.edu, Human Genome Release May 2004). GC content and information about the genomic position are also shown above. The different regions of the MHC are highlighted in different colours: classical class I – yellow; class III – green; classical class II – blue; and extended class II – red (according to classification in Horton et al 2004). Please note that certain genes appear more than once due to the presence of alternative splice variants.

complement system and cytokine molecules that are encoded in the MHC class III region, others have important functions in immunity, as discussed below.

Almost every individual is heterozygous for the maternally and paternally-inherited series of HLA alleles, also referred to as haplotypes, both of which are co-dominantly expressed. The high degree of polymorphism within the HLA genes and the existence of several genes that can perform the same function, result in a tremendous diversity of MHC molecules within the population. This is thought to have important implications in resistance of a host population to evolving pathogens.

1.3.1 Overview of the MHC molecules and their role in immunity

The MHC class I region encodes HLA-A, -B and -C, while the class II region encodes HLA-DP, -DQ and -DR. These are known as the classical molecules and are generally referred to as MHC class Ia and class IIa molecules respectively. Some of the genes encoding the non-classical molecules in each one of the two classes, also referred to as class Ib or class IIb molecules, have high sequence homology to the classical genes yet differ in function. Nonetheless, the non-classical molecules also have a role in immunity, as do the products of various other genes that reside within the class I and II regions.

1.3.2 Classical class I molecules: structure and function

The folding and assembly of a complete MHC class I molecule presenting a peptide antigen involves several steps. It requires a number of accessory proteins with chaperone-like function and takes place primarily in the endoplasmic reticulum (ER). The molecule is released from the ER and allowed to reach the cell surface only after a peptide has bound. MHC class I molecules present peptides derived from cytosolic proteins. The peptides are mainly generated by macromolecular structures called proteasomes. These consist of 28 subunits, some of which have proteolytic activity, such as PSMB8 and PSMB9 (Tanaka 1998). The pathway of MHC class I presentation is therefore often called the "cytosolic" or "endogenous pathway".

The classical MHC class I molecules are heterodimers, consisting of the α-chain, a ~44 KDa transmembrane heavy chain glycoprotein encoded by *HLA-A,-B* or *-C*, and a non-covalently bound 12 KDa soluble protein called β_2 microglobulin. The α-chain consists of three extracellular domains: α_1 and α_2 form the peptide binding cleft – with ability to bind one peptide 8-10 residues long, while the third compartment, α_3, is situated just underneath. This is followed by the hydrophobic transmembrane domain, and a cytoplasmic anchor segment (Bjorkman, Saper et al. 1987).

In order to bind to MHC class I molecules, proteasome-generated peptides are translocated from the cytosol to the ER lumen by the Transporter associated with Antigen Processing (TAP). TAP is a heterodimeric multimembrane-spanning polypeptide, consisting of two subunits, TAP1 and TAP2 (Kelly, Powis et al. 1992). They form a peptide binding site and two ATP binding sites that face the lumen of the cytosol. TAP binds peptides on the cytoplasmic site and translocates them to the lumen of the ER in an ATP-dependent step (Androlewicz, Anderson et al. 1993). The MHC class I molecule is then loaded with peptides in the lumen of the ER. The process involves TAP and several other molecules, such as the TAP-binding protein (TAPBP) (Ortmann, Copeman et al. 1997), the lectin chaperones calnexin and calreticulin, and the protein disulfide isomerase ERp57 (Hughes and Cresswell 1998; Oliver, Roderick et al. 1999). The fully folded MHC class I molecule and its bound peptide are transported to the cell surface through the secretory pathway. The N-glycan regions of the protein undergo several post-translational modifications in the ER and the Golgi apparatus (Parham 1996), during which the MHC molecule becomes fully matured.

1.3.3 Non-classical class I molecules

The non-classical molecules HLA-E, -F and -G are also involved in antigen presentation. However, these are not polymorphic as their classical counterparts. It has been suggested that they bind a restricted subset of peptides derived from the leader peptides of other class I molecules (Diehl, Munz et al. 1996; Braud, Jones et al. 1997; Lepin, Bastin et al. 2000). Furthermore, there is evidence that these molecules

contribute to the normal progression of pregnancy and growth of the developing foetus. This is achieved by inhibiting the cytotoxic activity of natural killer (NK) cells when these molecules are expressed on foetus-derived trophoblasts (cell type mediating the implantation of the foetus into the placenta) (Ishitani, Sageshima et al. 2003; Ishitani, Sageshima et al. 2006) perhaps by inhibiting NK migration across the endothelium (Dorling, Monk et al. 2000).

1.3.4 Other class I region encoded proteins with immune-related function

The protein products of the two class I-like genes, *MICA* and *MICB*, also have a role in NK cell-mediated immune response. Reports until now have shown that neither of the two molecules can bind with β_2 microglobulin nor do they present antigens, despite the high level amino acid homology (up to 30%) with classical class I genes (reviewed by Stephens 2001). The MIC molecules are recognised as ligands by the NKG2D receptor that is found on cells with NK activity (Stephens 2001). The signal is transmitted to the interior of the cell via DAP10, which interacts with and activates intracellular protein tyrosine kinases (Wu, Groh et al. 2002).

IER3 is a stress-inducible gene that is rapidly activated by viral infection and inflammatory cytokines such as the tumour necrosis factor (TNF). The promoter contains several consensus sequences for transcription factors including NF-κB. The protein has immunity-associated functions; it also contains domains that can confer a pro- or anti-apoptotic role (Taddeo, Esclatine et al. 2003; Shen, Guo et al. 2006).

1.3.5 MHC class II molecules: structure and function

The classical MHC class II molecules are also membrane-bound glycoproteins consisting of two homologous peptides: an α and a β chain. The non-covalently associating α and β chains, ~32 KDa and ~29 KDa respectively, have an extracellular part, a transmembrane segment and a cytoplasmic anchor. The external part of each polypeptide consists of two domains, i.e. α1:α2 and β1: β2, where the first of each

pair contributes to the formation of the antigenic-binding cleft (Stern, Brown et al. 1994).

The classical class II molecules present peptides derived from proteins that are either bound to the cell membrane or associated with vesicles of the endocytic processing pathway. Such proteins can become internalised by receptor-mediated endocytosis or phagocytosis (Honey and Rudensky 2003). The endocytic pathway appears to involve increasingly acidic endosome compartments, whereby the antigen successively moves down the pH gradient, eventually becoming degraded to about 13-18 residues. Lysosomes are membrane bound organelles which form at the Golgi complex and contain several types of acid-dependent hydrolases, such as cysteine proteases that are able to digest proteins (Honey and Rudensky 2003).

Classical class II molecules are synthesised within the ER where they associate the invariant (Ii) chain. This trimeric protein interacts with the peptide binding pocket of the class II molecule thereby hindering binding of endogenously derived peptides, which are predestined as substrates for class I molecules (Cresswell 1996). The complexes are transported from the ER and via the Golgi complex to early endosomes. As they progress into more acidic compartments with increasing proteolytic activity, the Ii chain is gradually degraded with only a small fragment (class II associated invariant chain peptide - CLIP) remaining within the peptide binding site of the class IIa molecule (Bakke and Dobberstein 1990). HLA-DM is a class IIb molecule that facilitates the exchange of CLIP with antigenic peptides and endorses strongly-binding ones, an activity that is inhibited in the presence of HLA-DO, which is also a class IIb molecule. The MHC class II-peptide complexes are subsequently presented on the cell membrane (Ting and Trowsdale 2002; Busch, Rinderknecht et al. 2005).

1.3.6 Other class II encoded proteins with immune-related function

Another MHC class II encoded protein with function in immunity is RXRβ, a member of the retinoid X receptor family, which is involved in mediating the effects of retinoic acid (RA). RXR molecules form homodimers and heterodimers with other

receptors such as the RA receptor (RAR). After ligand-binding, the dimer targets DNA sequences corresponding to hormone response elements (HRE) to activate transcription of the relevant genes (Szanto, Narkar et al. 2004). RXR molecules have also been implicated in myeloid cell differentiation (Collins 2002).

The polypeptide encoded by the DAXX gene is a death-domain protein which has a role in apoptotic cell death. However, its localisation in both the nucleus and the cytoplasm suggests multiple functions, including a possible role in cell survival, although the precise mechanism is not clear (Salomoni and Khelifi 2006). According to recent reports, it is involved in growth arrest and/or apoptosis in early B-cell development (Muromoto, Sugiyama et al. 2004).

1.3.7 The role of MHC class III encoded molecules in the complement system

The complement system is the major defence effector of the humoral branch of the immune system. Several proteins are implicated in the various pathways of highly regulated cascades that constitute this system, the main functions of which include: i) lysis of viruses, bacteria and cells; ii) opsonisation (promotes phagocytosis of loose antigen particles); iii) binding to specialised receptors on leukocytes to trigger activation of immune responses (e.g. inflammation and secretion of immunoregulatory molecules); and vi) immune clearance (removal of circulating immune complexes to liver and spleen). Many components are proenzymes; they are inactive until proteolytic cleavage removes an inhibitory fragment to expose the active site.

The early steps in complement activation that culminate in the formation of C5b, a central component of the lysis-inducing membrane attack complex (MAC), can occur via one of three pathways: the classical, the alternative, and the lectin pathway. The classical pathway, part of the adaptive immunity, is activated by antibody bound to foreign material. On the other hand, the alternative pathway which involves differential interaction of a range of inhibitors and cofactors with self and non-self surface structures, and the lectin pathway which recognises foreign carbohydrate

structures, can be activated innately. [See Appendix B for more information on the complement system]

Three complement genes reside within class III of the MHC. The *C4* gene, which is polymorphic – *C4A* and *C4B* – (Blanchong, Chung et al. 2001), is transcribed and translated to a single-chain precursor of ~192 KDa. The precursor is cleaved to yield chains α, β and γ in a disulfide-linked trimer, prior to secretion from the cell into blood circulation. Activation of secreted C4 is achieved when the α chain is cleaved by C1 into C4a and C4b, where the latter remains in complex with the β and γ chains. Further degradation of C4b by C1 into the inactive fragments C4c and C4d blocks the generation of C3 convertase (www.genecards.org/cgi-bin/carddisp.pl?gene=C4A and www.genecards.org/cgi-bin/carddisp.pl?gene=C4B).

The other MHC class III classical pathway component, C2, is a secreted protein of ~83 KDa. When activated, it can selectively cleave the Arginine - Serine bond in the C3 α chain to form C3a and C3b, and the Arginine - Xaa (i.e. any aminoacid) bond in complement component C5 α chain to form C5a and C5b (www.genecards.org/cgi-bin/carddisp.pl?gene=C2).

Finally, the alternative pathway molecule Factor B, circulates in the blood as a single polypeptide of ~86 KDa. The larger fraction, Bb, has similar proteolytic activity as C2a, and has also been implicated in various other processes such as the proliferation and differentiation of pre-activated B lymphocytes, the rapid spreading of peripheral blood monocytes, the stimulation of lymphocyte blastogenesis and the lysis of erythrocytes (www.genecards.org/cgi-bin/carddisp.pl?gene=BF).

1.3.8 Other class III encoded proteins with immune-related function

Three genes: *HSPA1A*, *HSPA1B* and *HSPA1L*, encoding 70 KDa heat shock proteins with immune-related function, are present in the class III region. These heat shock proteins recognise non-native protein-conformations, and cooperate with other chaperones to stabilise proteins and prevent formation of aggregates. They also mediate the folding of newly translated polypeptides in different cell compartments,

and bind hydrophobic peptide segments that become exposed i) during translation or translocation through the cell membrane, or ii) subsequent to stress-induced damage. There is mounting evidence that members of the heat shock protein (HSP) family have a role in immunity (Gullo and Teoh 2004). They reportedly interact with antigen presenting cells through receptors like CD91 to elicit various cascades of events. As a result, HSP-chaperoned peptides are presented by MHC molecules, NFκB translocates to the cell nucleus, and dendritic cells mature; all these effects contributing to activation of antigen presenting cells (Li, Menoret et al. 2002; Srivastava 2002; Srivastava 2002).

The tumour necrosis factor (TNF) cluster has three genes encoding precursors of the cytokines: lymphotoxin alpha (LTα), lymphotoxin beta (LTβ), and TNF. These molecules mediate various inflammatory, immunostimulatory and antiviral responses. They have also been implicated in lymphoid development (reviewed in Locksley, Killeen et al. 2001; Eissner, Kolch et al. 2004).

Other MHC class III genes suggested to have immune-related functions include: *AIF1* with a role in inflammation (Deininger, Meyermann et al. 2002); *NCR3* with a role in activation of cytotoxicity in NK cells (Pende, Parolini et al. 1999); and *LST1* with a role in lymphocyte proliferation (Rollinger-Holzinger, Eibl et al. 2000). Furthermore, a cluster of genes that are likely to encode proteins belonging to the lymphocyte antigen-6 family is also present in the class III region. Members of this family are cell surface proteins with a glycosyl-phosphatidyl-inositol domain. Although their precise function is has not yet been elucidated, there is evidence to suggest that they are involved in immunity (Mallya, Campbell et al. 2002).

1.3.9 Early stages in the classical MHC class II gene induction: the IFN-γ signalling pathway

The presence within a 4 Mb genomic segment of such a large number of genes with related biological function, several of which participate in identical pathways or mechanisms, advocates a biological significance for this clustering. In fact, expression of several of the MHC genes which have a role in antigen presentation is coordinately

regulated by various downstream effectors of cytokine signalling pathways, such as interferon gamma (IFN-γ).

IFN-γ is secreted by activated T-cells and NK cells, and among others, it is involved in antiviral responses and immune surveillance. IFN-γ interacts with a specific receptor composed of subunits IFNGR1 and IFNGR2, which are expressed on almost all cell types. Upon binding of IFN-γ, the receptor oligomerises, and activates Jak1 and Jak2 by phosphorylation. Subsequently, the activated Jaks phosphorylate tyrosine 440 on the intracellular domain of IFNGR1, thus enabling Stat1 binding. Additional tyrosine residues become phosphorylated on the receptor and the Jaks, as is tyrosine 701 on Stat1, which forms dimers. Stat1 is subjected to another phosphorylation event at serine 727 and translocates to the nucleus. Once inside the nucleus, the dimer regulates gene expression by binding to γ-activated sequence (GAS) elements at gene promoters (reviewed by Stark, Kerr et al. 1998). CBP/p300, MCM5 and BRCA1, were identified as co-activators of Stat1 (Zhang, Vinkemeier et al. 1996; Zhang, Zhao et al. 1998; Ouchi, Lee et al. 2000). Among the wide range of genes induced by IFN-γ via Stat1 are the interferon response factor 1 (IRF1) and the MHC class II transactivator (CIITA), which have a role in upregulating the expression of several MHC genes, as well as of *hsp70* genes (Stephanou, Isenberg et al. 1999; Stephanou and Latchman 1999), in the class III region.

1.3.10 Transcriptional regulation of MHC genes: the class II transactivator

The class II transactivator (CIITA) has been referred to as the master regulator for the expression of the MHC class II genes. Constitutive expression of the class II molecules is largely restricted to antigen presenting cells (APC): monocytes, dendritic cells, B cells and macrophages. Activated human T cells and thymic epithelial cells also express class II genes. Various stimuli can induce expression of class II genes, IFN-γ being the most important, while others have an inhibitory effect.

Expression of MHC class II genes is primarily regulated at the level of transcription. The promoters of these and related genes are characterised by the presence of

conserved sequence elements, the: W/S, X (comprising X1 and X2), and Y boxes. Factors that have been shown to bind to these sequences are the trimeric complex RFX, the X2BP complex, and another trimeric complex NFY. Combined interaction of these constitutively and ubiquitously-expressed factors with the regulatory elements of MHC class II promoters is necessary for gene expression. CIITA, a non-DNA-binding co-activator, sits on the macromolecular complex that is formed (reviewed by Ting and Trowsdale 2002). DO expression is mainly restricted to B cells, where HLA-DOA is activated by CIITA. Although expression of HLA-DOB can be induced by CIITA, the precise regulation mechanism for this gene remains unclear (Nagarajan, Lochamy et al. 2002).

CIITA also contributes to the expression of classical and non-classical MHC class I genes, though to a lesser extent. Some studies have suggested that it might have other roles, both within and outside the immune system (Nagarajan, Lochamy et al. 2002). Nonetheless, the MHC class II and their related genes, remain the major targets of CIITA.

Repression of CIITA has been shown to result in downregulation of MHC class II expression. This is a process that occurs naturally during dendritic cell maturation (Landmann, Muhlethaler-Mottet et al. 2001) as well as during terminal differentiation of B cells into plasma cells (Piskurich, Lin et al. 2000). Furthermore, both constitutive or IFN-γ induced expression of CIITA was reported to be silenced in fetal trophoblasts, where different mechanisms have been proposed (Geirsson, Paliwal et al. 2003).

1.3.11 Sequence features within the MHC region

The heterogeneity in isochore structure that has been described throughout the human genome also marks the sequence within the MHC. The classical class II region has an average GC content of ~40%, and thus classifies as an L2 isochore (Stephens, Horton et al. 1999), whereas the neighbouring class III, with ~53% GC, is an H3 isochore. Centromeric to the classical class II is the extended class II region which has a sharp transition in GC content to ~51%, and is therefore an H2 isochore. Finally, the class I

region has a GC content of intermediate value, with an average of ~46% GC and classifies as an H1 isochore. Features such as the GC-richness of the sequence have been examined experimentally in order to identify the important underlying characteristics and to allow analyses from an evolutionary and other perspectives (Pavlicek, Clay et al. 2002).

At the class II/III boundary, close to the point where a steep transition in GC content exists, there is high sequence homology (~80% identity) to the pseudoautosomal boundary, the sequence separating the sex-specific and pseudoautosomal regions on human sex chromosomes (Fukagawa, Sugaya et al. 1995). Although this pseudoautosomal boundary-like sequence (PABL) is not at the precise point of GC% transition, it has been proposed that it is characteristic of isochore junctions or chromosome band boundaries. Several PABL sequences have been identified across the genome (Fukagawa, Nakamura et al. 1996), but these are not present at every single boundary. The boundary between the classical and the extended class II, despite the sharp transition in GC content, does not contain a PABL sequence (Stephens, Horton et al. 1999).

The relative AT-richness of the classical class II region suggests that it localises to a G band. Senger *et al.* have shown that class II FISH probes mapped to the R band 6p21.31, on metaphase chromosomes (Senger, Ragoussis et al. 1993). A high-resolution prometaphase G banding, however, has identified two thin G-subbands within 6p21.31 (Yunis 1981), one of which could correspond to the AT-rich classical class II region. Interestingly, a recent analysis based on mathematical modelling has defined the specific base position of the isochore transition (Zhang and Zhang 2003).

The MHC also houses several repetitive elements like SINEs and LINEs (Andersson, Svensson et al. 1998). The classical class II region is a model for studying the influence of retroelements on genomic activity, plasticity and evolution (Beck and Trowsdale 2000). Repeat elements make up about 23% of the class II region (Andersson, Svensson et al. 1998) and contribute to its polymorphism.

1.4 Aim of this thesis

The association of replication timing with fixed genomic features, on the one hand, and with variable transcription or chromatin conformation, on the other, poses a problem for our understanding of the organisation of replication across the genome. The patterns and sizes of replication bands in metaphase chromosomes do not fluctuate in different cell types. Yet studies at individual genomic locations usually show variations in replication timing according to transcriptional activity. Which of these factors are the most significant correlates of replication timing? Does replication timing change following transcriptional induction of coordinately regulated clusters of genes in larger genomic regions? What is the replication time in a genomic region that has a combination of features that individually would be associated with either early or late replication?

I have approached these questions by investigating the organisation of DNA replication in different cell types in the Major Histocompatibility Complex (MHC), which contains clusters of co-regulated and inducible genes positioned within distinct isochores. This aim was addressed in three different chapters with the following objectives:

1. To assess replication timing across ~1.5 Mb of the human MHC using a suitable biochemical assay in three cell lines of different tissue derivation. The data were used to address the relationship of replication timing with expression status and sequence organisation. – Chapter 3.

2. To assess replication timing for a lymphoblastoid cell line across ~3.8 Mb of the human MHC using a tile-path array of 2.5 Kb clones. The data were used to address the relationship between replication timing and sequence organisation such as gene density, GC content, and abundance of repetitive sequences. – Chapter 4.

3. To identify potential sites of DNA replication initiation using nascent DNA. A custom tile-path array covering the human MHC was used for this analysis.

Several candidate regions were examined by PCR to verify the findings of the array. The relationship between nascent DNA enrichment and sequence organisation, such as CpG islands and exon density, was also examined. Existing theories concerning origins of replication, such as zones and sites of initiation were addressed. – Chapter 5.

Chapter 2

Materials and Methods

2.1 Cell culture

Cell lines were grown on ungelatinised tissue culture plasticware in cell culture incubators that maintained a temperature of 37 °C and a humidified atmosphere complemented with either 5% or 10% CO_2. Growth media and other cell culture reagents were obtained from Cancer Research UK cell services (Clare Hall).

AHB is a human EBV-transformed B-lymphoblastoid cell line which retains two normal copies of chromosome 6 and originates from a female individual. It was kindly provided to our laboratory by the late Lady Julia Bodmer (Oxford, UK). It was grown in RPMI-1640 media supplemented with 10% Fetal Calf Serum (FCS), 2mM L-glutamine, 100 units/ml penicillin V and 0.1 mg/ml streptomycin.

PGF is a human EBV-transformed B-lymphoblastoid cell line (ECACC 94050342) derived from a consanguineous male and is homozygous for the MHC. Similar to AHB, PGF cells were also grown in RPMI-1640 which was supplemented as above.

HL60 is a human myeloid leukaemia cell line (ATCC CCL-240) derived from a female. It was grown in Dulbecco's Modified Eagle's Medium (DMEM) with 10% FCS and other supplements as described above.

AHB, HL60 and PGF cell lines were grown in suspension and were maintained at exponential growth by addition of fresh media every 2-3 days such that their concentration would not exceed 10^6 cells/ml.

MRC5 cells are human male fetal lung fibroblasts with a normal diploid karyotype (ECACC 97112601). They were cultured as monolayers in RPMI-1640 supplemented as above, and were passaged when they reached confluence, usually every 3-4 days.

Cells were washed with phosphate buffered saline (PBS) and were then detached with 4 ml of a solution containing porcine trypsin and EDTA per 75 cm^2 culture flask. After 5 minute incubation all the cells became detached and the trypsin digestion was terminated by the addition of an excess of culture medium containing FCS. Cells were then reseeded in two to three new flasks at 30-50% confluence.

Chinese hamster ovary (CHO) cells, a subclone of the parental CHO cell line that was originated by Puck (Puck, Cieciura et al. 1958), were grown as monolayers in Alpha medium supplemented with 10% FCS and 2 mM L-proline. When they reached confluence, they were split and reseeded at 30% confluence.

2.2 Aphidicolin synchronisation

HL60 cells were incubated at a concentration of $5x10^5$ cells/ml with 1μg/ml aphidicolin (stock at 10mg/ml in dimethyl sulphoxide), for 24 hours to arrest cells before entering into S-phase. To release from the block, cells were twice pelleted (by centrifugation in a conical-bottomed tube at 1000 rpm for 5 minutes in swinging bucket rotor) and resuspended in fresh culture media, before finally resuspending in fresh culture media at a concentration of $5x10^5$ cells/ml. They were then incubated for 12 hours before further addition of aphidicolin to the media at 5 μg/ml for 12 hours after which the cells were released from the block as described above. A fraction of cells was removed from the total population at one hour intervals from the time of release and labelled with bromodeoxyuridine (BrdU) prior to fixation.

2.3 Labelling DNA with Bromodeoxyuridine

BrdU is a halogenated nucleotide that is incorporated instead of thymidine during DNA synthesis. To label newly-synthesised nascent DNA, AHB, HL60, MRC5 cells were incubated with 100 μM BrdU for 15 minutes (confocal microscopy) or 1 hour (BrdU immunoprecipitation) before fixation. CHO cells on the other hand were labelled with BrdU for 48 hours. For fixation the cells were harvested as appropriate and pelleted. The supernatant was removed and the cell pellet was vortexed gently

while ice-cold 70% ethanol was added dropwise. Fixed cells were kept in 70% ethanol on ice for at least 30 minutes prior to any further treatment.

2.4 Immunostaining for flow-cytometry

To test for expression of the MHC class II molecule HLA-DRA, cells were incubated with an anti-HLA-DRA antibody and analysed by flow cytometry. Ten million (10^7) cells were pelleted and washed in PBS and then resuspended in 1 ml PBS. The suspension was aliquoted in two FACS tubes and in the one, FITC-conjugated anti-HLA-DR antibody (347363, Becton Dickinson) was added at final concentration of 25 μg/ml, while the other was kept as negative control. Both tubes were left on ice for 20 minutes and afterwards were made up to 10 ml with PBS and pelleted again. They were resuspended in 0.5 ml PBS and fixed by adding 0.5 ml 1% v/v formaldehyde solution.

2.5 Cell cycle analysis or sorting by flow-cytometry

To prepare cells for immunofluorescence and FACS analysis or sorting, ethanol fixed cells were washed twice in PBS.

For BrdU immunolabelling, cells were resuspended in 2 M HCl and incubated at room temperature for 30 minutes with mixing at regular intervals. They were then centrifuged and twice washed in PBS and once in PBST (0.2% v/v Tween 20 in PBS) with 0.1% w/v BSA to remove the acid. Mouse monoclonal BrdU antibody (Becton Dickinson) was added directly to the cell pellet, 2 μl per 10^6 cells, and the cell suspension incubated at room temperature for 20 minutes followed by two washes in PBST with 0.1% w/v BSA. Cells were stained with 50 μl FITC-conjugated polyclonal rabbit anti-mouse immunoglobulin $F(ab')_2$ fragments (DAKO) (at 1 in 10 dilution from stock) for 20 minutes at room temperature in the dark and then washed in PBS.

To assess DNA content, cells were treated for 15 minutes with 100 μg/ml of ribonuclease A (Sigma) and finally resuspended in 50 μg/ml propidium iodide

solution (PI) for at least 10 minutes before analysis. After preparation samples were kept on ice. Cells were sorted on a FACS Vantage Cell Sorter (Becton Dickinson) or analysed on a FACS Calibur (Becton Dickinson) for their DNA content. PI is an analogue of ethidium bromide that intercalates with RNA and DNA molecules. The binding markedly enhances the fluorescence of the dye upon excitation.

Fluorochromes were excited by a 488 nm laser. FITC fluorescence was collected between 515 and 545 nm and PI fluorescence between 650 and 690 nm. Data were collected with the associated software of the equipment used and further analysis performed with the programme CellQuest (Becton Dickinson) and FloJo (Tree Star). Flow-cytometry experiments and data analysis were performed with the help of Dr Gary Warnes and Dr Ayad Eddaoudi of the FACS laboratory, at Cancer Research UK London Research Institute.

2.6 Confocal microscopy

Centrifuge tubes containing 0.5 ml PBS were used to collect 10^5 cells from each S-phase fraction. The cell suspension was overlaid on poly-L-lysine coated coverslips placed in custom made centrifuge adaptors. Following centrifugation at 1000 rpm in a swinging bucket rotor for 10 minutes, the coverslips were inverted onto 10 µl drops of Mowiol mountant on glass microscope slides. Coverslip mounted slides were left to set overnight at room temperature and then stored at 4 °C in the dark. Assistance and materials were kindly provided by Alastair Nicol and Peter Jordan of the Light Microscopy Laboratory, Cancer Research UK London Research Institute.

Slides were examined with a Zeiss LSM 510 confocal laser scanning microscope equipped with a Plan Apochromat 63x/NA 1.4 objective. Excitation of the fluorescent dyes PI (red) and FITC (green) was performed using the 488 nm Argon laser. Optical section images were collected at 0.4 µm z-intervals through each nucleus using the LSM 510 imaging software.

2.7 Isolation of BrdU-enriched DNA

For immunoseparation of BrdU labelled DNA, FACS sorted cells were collected in a 1.5 ml tube containing 400 µl of lysis buffer (1 M NaCl, 10 mM EDTA, 50 mM Tris-HCl [pH 8.0], 0.5% SDS, 0.2 mg/ml proteinase K, 0.25 mg/ml salmon DNA) and incubated for 2 hours at 50 °C. At this stage the samples could be stored at -20 °C for future experiments. BrdU-labelled Chinese hamster DNA was added at 5 ng per 10^4 cells to each sample to enable assessment of the immuno-precipitation efficiency at a later stage.

DNA was extracted with a standard Phenol Chloroform Isoamyl Alcohol (IAA) (25:24:1) (Invitrogen) extraction in Phase Lock Gel (PLG) heavy 2 ml tubes (Eppendorf), as follows. Each sample was transferred into a 2 ml tube with heavy PLG and 500 µl of Phenol Chloroform IAA was added. The tubes were sealed and shaken vigorously for at least one minute. They were then placed in a microcentrifuge and spun at 13000 revolutions per minute (rpm) for 5 minutes. (Note: During centrifugation the gel forms a boundary between the aqueous phase which is conveniently situated above the gel, while the interphase and the organic phase remain below.) The DNA containing aqueous phase was collected from each sample and transferred to new 1.5 ml tubes to which 1 volume of Isopropanol was added and mixed thoroughly by shaking. The samples were then left overnight at -20 °C followed by centrifugation at 4 °C with 15000 rpm for 30 minutes to precipitate the DNA. The pellets were washed in 1 ml of 70% ethanol and air-dried before dissolving in 480 µl of TE buffer and adding 20 µl of sheared & denatured salmon or herring DNA (Sigma) (10 mg/ml).

To shear the DNA, samples were sonicated for 20 seconds at 21% amplitude (yielding fragments with an average size of about 700 bp) using a sonicator mounted with a 2 mm diameter probe (Vibracell, Sonics) and kept on ice. Then they were denatured for 4 minutes at 95 °C and snap-cooled on ice. One by one the samples were adjusted with 100 µl pre-IP stock buffer (to 10 mM sodium phosphate [pH 7.0], 0.14 M NaCl, 0.05% Triton X-100) before adding 2 µg (80 µl at 25 mg/ml) of anti-BrdU monoclonal antibody (Becton Dickinson Immunocytometry). The samples were then

incubated at room temperature for 25 minutes with end over end rotation. After brief centrifugation to draw the contents to the bottom of the tube, the antibody-bound BrdU-enriched DNA in each sample was precipitated by adding 35 μg of rabbit IgG [directed against mouse IgG (Sigma)] and incubating at room temperature for 25 minutes with constant rotation. Samples were centrifuged at 4 °C with 15000 rpm for 15 minutes. The supernatant was carefully removed with needle and syringe to avoid disturbing the pellet. The pellet was then washed in 750 μl of ice cold Washing buffer (10 mM sodium phosphate [pH 7.0], 0.14 M NaCl, 0.05% Triton X-100) and placed on a vortex for 30 seconds. To digest the proteins, the samples were then resuspended in 200 μl of Digestion buffer (50 mM Tris-HCl [pH 8.0], 10 mM EDTA [pH 8.0], 0.5% w/v Sodium Dodecyl Sulphate (SDS), 0.25 mg/ml Proteinase K) and were incubated overnight at 37 °C. An additional 100 μl of this buffer was then added and the tube incubated for 1 hour at 50 °C. DNA was extracted with 300 μl of Phenol Chloroform IAA in PLG light 2 ml tubes (Eppendorf), as described above. The aqueous phase was recovered and 20 ng of yeast tRNA (Sigma) was added (4 μl of 5 μg/ml stock) to act as a carrier. The DNA was precipitated with 3 volumes absolute ethanol and 0.1 volume sodium acetate followed by a similar procedure to that described above. The DNA pellet was dissolved in TE buffer at a concentration of 1000 cell equivalents per μl and stored at 4 °C for up to 2 months.

2.8 Polymerase Chain Reaction – conventional

Primer pairs were designed using the open access software "Primer 3" (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_ www_slow.cgi) with the following parameters: sequence input range was 400 to 800 bp; regions with repetitive sequences – as defined by the human genome sequence database at http://genome.ucsc.edu/cgi-bin/hgTracks using the "get DNA" function – were excluded from primer selection; amplification product size requirements were fulfilled by minimum 200 bp, optimum 300 bp and maximum 400 bp; primer size requirements were fulfilled by minimum 20 bp, optimum 22 bp and maximum 24 bp. All other parameters were left at the default values.

PCR reactions were prepared in 50 µl reaction volumes with 25 µl of Qiagen Taq PCR Mastermix, 5 µl of forward and reverse primer mix at 5 M each (500 nM final concentration), DNA template empirically defined as per requirements, and the remaining volume filled with distilled water. Cycling conditions were at 95 °C for 2 minutes, followed by cycles of 95 °C for 35 seconds, 55 °C for 30 seconds, 72 °C for 30 seconds and final extension at 72 °C for 5 minutes. These reactions were performed in a DNA Engine (MJ Research) thermal cycler.

2.9 Agarose gel electrophoresis

To visualise and establish the size of nucleic acids after the various treatments and experimental procedures, agarose gel electrophoresis was performed. Agarose was prepared by boiling a mixture of agarose powder (Gibco) in 1 x TAE at concentrations of 1-2% w/v depending on the expected size of the DNA fragment. After cooling to 55 °C, ethidium bromide was added to a final concentration of 0.25 µg/ml so that the DNA could be visualised under UV light. A gel-casting tray with a comb was used to set molten agarose into a gel and after solidifying was placed in a running tank and covered by 1 x TAE. Samples with additional tracking dye – (Orange-G, Sigma) at 1:6 dilution – were loaded according to the volume available in the well, in the range of 10-35 µl. To allow sizing of the fragments 5 µl of 1 Kb DNA ladder (Invitrogen) were also loaded on either side of the samples. The gel was then electrophoresed at 100-140 V for 15-60 minutes depending on the separation required, and photographed using a UV trans-illuminator (260nm) and camera.

2.10 Band Excision and DNA extraction

PCR products were purified using the QIAquick gel extraction kit (Qiagen) according to the manufacturer's protocol. This involved solubilising the portion of the agarose gel containing the required band, binding of the DNA to a column, followed by a series of washes and elution of the DNA.

2.11 Direct sequencing of PCR products and analysis

The sequencing reaction was performed using the Big Dye Terminator reaction kit (Applied Biosystems) according to manufacturer's instructions, with subsequent analysis on the 3730 DNA Analyzer (Applied Biosystems).

Sequences were examined using the open access 4 Peaks software, available at (http://mekentosj.com/4peaks/). Human genome database searches using BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat) were performed to ensure that the correct region of DNA had been amplified.

2.12 Agarose gel treatments, Southern blotting and hybridisation

The BrdU-enriched nascent DNA samples were first PCR-amplified followed by agarose gel electrophoresis. The gel was then rinsed in distilled water before submerging in denaturation buffer (1.5 M NaCl, 0.5 M NaOH) and gently agitating for 30 minutes. After rinsing in distilled water the gel was submerged and gently agitated in neutralisation buffer (1.5 M NaCl, 0.5M Tris-HCl [pH 7.5]) for 30 minutes. Both steps were carried out at room temperature. The gel was rinsed with distilled water and transferred overnight to Hybond N+ positively charged nylon membrane (Amersham) according to the manufacturer's protocol. The DNA was immobilised onto the membrane by crosslinking in a UV crosslinker (Stratagene). The membrane was then soaked in 2x SSC and pre-hybridised in a rolling hybridisation tube with Rapid-Hyb buffer (Amersham) – 15-20 ml per 20 cm^2 – at 65 °C for 1 hour in a suitable incubator (Techne).

Meanwhile, using the Prime-It II Random Labelling kit (Stratagene), 25 µl at 1 ng/µl of the equivalent DNA probe, extracted from gel as described above, together with 10 µl random oligonucleotide primers was denatured at 95 °C for 5 minutes and immediately placed on ice. The contents were drawn to the bottom of the tube with brief centrifugation and 10 µl of 5x dCTP buffer, 5 µl of [α-^{32}P]-dCTP (50 µCi)

(Amersham), and 1 µl of Exo(-) Klenow (5 U/µl) were added in this order. Gentle pipetting was used to mix the labelling reactants, followed by incubation at 37 °C for 10 minutes. The reaction was stopped by adding 2 µl stop mix (0.5 M EDTA, pH 8.0).

Unincorporated nucleotides were removed using MicroSpin™ G-50 Columns (Amersham) as recommended by the manufacturer; the labelled probe was then denatured at 95 °C for 5 minutes and then placed on ice for 2 minutes. The contents were briefly centrifuged and the labelled probe was mixed with the buffer in the hybridisation tube. Hybridisation was carried out overnight at 65 °C with rolling of the tube. The membrane was washed twice for 10 minutes with low stringency buffer (2x SSC, 0.1% w/v SDS) and then twice for 10 minutes with high stringency buffer (1x SSC, 0.1% w/v SDS), both of which had been pre-warmed at 65 °C. The membrane was exposed to film (Amersham) for up to 1 hour and then developed and also to a screen for Phosphorimager (Molecular Devices) scanning. Phosphorimager data were analysed using ImageQuant software.

2.13 RNA extraction from cultured cells

RNA was extracted using the TRIzol®Reagent (Invitrogen) following manufacturer's protocol with minor amendments. Briefly, 4.5 ml Trizol was added to a 75 cm^2 flask of MRC5 cells at 70% confluence, or 10^7 AHB cells. The mixture was pipetted repeatedly to allow cell lysis and the homogenised samples were incubated at room temperature for 5 minutes. 1 ml of chloroform was added to each sample, followed by vigorous shaking for 30 seconds and incubation for a further 3 minutes at room temperature. The samples were then centrifuged at 12000g for 5 minutes at 4 °C and the aqueous phase from each sample transferred to a fresh centrifuge tube. The RNA was precipitated by addition of 2.4 ml of isopropanol, incubation at room temperature for 10 minutes, and centrifugation at 12000g for 10 minutes at 4 °C. The RNA pellet was washed twice with 4.5 ml of 75% ethanol (in RNase-free water) by vortexing and then centrifuging for 5 minutes at 8000g at 4 °C. The RNA pellet was air-dried at room temperature and re-suspended in 100 µl RNase-free water. The quality of the

RNA was examined by agarose gel electrophoresis, it was quantified using the Nanodrop® ND-1000 spectrophotometer, and was then stored at -70 °C.

2.14 RNA reverse transcription

Complementary DNA synthesis was performed using the First Strand cDNA Synthesis kit (Amersham). Approximately 2.5 μg of total RNA was added up to 10 μl, total volume, of reaction mix which included 2 μl of random-hexamer primers (50 ng/μl) and 1 μl 10 mM dNTPs, and incubated at 65 °C for 5 minutes then placed on ice for 2 minutes. Nine μl of a mixture (consisting of: 2 μl of 10x RT buffer, 4 μl of 25 mM magnesium chloride, 2 μl of 0.1 M DTT solution and 1 μl of RNase inhibitor) were added to each reaction and mixed gently, then incubated at 25 °C for 2 minutes. One μl of Superscript™ II RT was mixed and the reaction was incubated at 25 °C for 10 minutes, followed by incubation at 42 °C for 50 minutes. Reactions were terminated by heating at 70 °C for 15 minutes and then chilling on ice. Control reactions were performed simultaneously, without the addition of enzyme. The cDNA product was used in real-time PCR experiments.

2.15 Isolation of leading strand nascent DNA

Asynchronous AHB and PGF cell cultures, which were maintained at exponential growth by splitting the day before, were used for extraction of nuclei. 1.5 x 10^8 cells were first washed once in ice-cold PBS buffer, and once in RBS buffer (10 mM Tris-HCl [pH 8.0], 10 mM NaCl and 3 mM $MgCl_2$). They were resuspended to 2.5 x 10^7 cells/ml by continuous vortexing and dropwise addition of RBS buffer. They were left on ice for 5 minutes and an equal volume of RBS containing 0.4% v/v non-ionic detergent NP-40 (Sigma) added, followed by incubation on ice for an additional 10 minutes. The nuclei obtained had to be handled with care to avoid lysis. They were pelleted by centrifugation for 10 minutes at 2000 rpm in a swinging bucket rotor, and washed in RBS buffer. They were then resuspended to 5 x 10^7 cells/ml in RBS buffer and an equal volume of lysis buffer (20 mM Tris-HCl [pH 8.0], 20 mM EDTA [pH

8.0], 2% w/v SDS, and 500 μg/ml proteinase K) added for overnight incubation at 56 °C.

Samples were transferred to 15 ml tubes containing PLG heavy (Eppendorf) and 5 ml of Phenol Chloroform IAA [25:24:1 (Invitrogen)] was added, followed by slow rotation for 10 minutes to allow proper mixing of phases. The samples were then centrifuged for 5 minutes at 3000 rpm in a swinging bucket rotor, and the aqueous phase was decanted in a glass beaker containing cold absolute ethanol. After the supernatant had been carefully removed, the DNA which had precipitated and settled at the bottom of the beaker was washed with ice cold 70% ethanol to rid of excess salt. The contents of the glass beaker were carefully transferred to a 50 ml Falcon tube and centrifuged at 3500 rpm in a swinging bucket rotor for 5 minutes. The high molecular weight DNA that remained was recovered after careful removal of the supernatant with syringe and needle. The DNA was dissolved in 1 ml distilled water with overnight incubation at 65 °C with gentle shaking.

Centrifugation tubes (Beckman) containing 5-20% w/v sucrose gradients were prepared for isolation of DNA with particular buoyant density. Sucrose solutions of 20, 17.5, 15, 12.5, 10, 7.5, 5 and 2.5% w/v were prepared and 5ml of each was added in the same order from the bottom of the tube up with care to avoid mixing between the different layers of sucrose solution. Three hundred μg of genomic DNA were denatured by boiling at 95 °C for 5 minutes and cooled on ice. It was then applied to the top of each gradient. One hundred μg of double-stranded 1 Kb DNA ladder (Invitrogen) were loaded in exactly the same way on a parallel gradient. Samples were centrifuged for 20 hours at 20 °C in a Beckman SW28 rotor at 26000 rpm. Thirty-four 1 ml fractions were collected for every sample in 1.5 ml tubes with extreme care to avoid mixing of the linear sucrose gradient that had formed. The fractions collected from the gradient of the 1 Kb ladder were analysed by agarose gel electrophoresis. Using the separation pattern of the marker as sedimentation velocity reference, fractions containing ssDNA ranging in size between 1-2 Kb were pooled together. DNA was recovered by ethanol precipitation and resuspended in 250 μl TE [pH 8.0] buffer. Nascent ssDNA was quantified using the Nanodrop® ND-1000

spectrophotometer, and was then stored at -20 °C. It was subsequently used in real-time PCR experiments and array hybridisations.

2.16 Polymerase Chain Reaction – quantitative

Two different types of quantitative PCR reaction are most commonly used nowadays. The first is referred to as TaqMan® PCR assay, and the other is known as SYBR Green real time PCR. Both are designed to monitor the process of amplification throughout the PCR reaction.

SYBR Green is believed to bind in the minor groove of double-stranded DNA and is fluorescent when bound. This binding characteristic is harnessed during PCR to monitor the process of amplification as PCR product is generated. An increasing amount of PCR product will result in an increase in SYBR Green dye fluorescence. This quantitative PCR assay was used extensively during this project. Primers were designed using similar methods and parameters to those for conventional PCR (described in section 2.8). Reactions were performed in a Chromo 4™ (MJ Research) thermal cycler, with initial denaturation for 5 minutes at 95 °C and 40 cycles of denaturation at 95 °C, annealing at 55 °C (unless otherwise stated) and extension at 72 °C, followed by a final 5 minute extension step at 72 °C. This was followed by sequential increase in temperature from 72 °C to 95 °C, to establish the melting temperature of the product. Analysis of the results was performed using different versions of the Opticon (MJ Research) amplification detection software.

2.17 MHC tile-path array

As a result of collaboration between Dr Denise Sheer and Dr Stephan Beck (Sanger Institute, Cambridge), we had access to a tiling path array covering ~3.8 Mb of the human MHC, between the genes *MOG* to *ZBTB9* (chromosome 6, base position 29739385-33552581). The array consisted of 9120 features, 6648 of which had been assigned a start and end position within the MHC. Each one of the 6648 features represented a ssDNA molecule. Notably, two ssDNA molecules for the forward and

another two for the reverse strand provided quadruplicate representation for every tile. The total number of individual tiles was 1662, and the average length of each tile was ~2.5 Kb. Around 99% of this region was encompassed by clones overlapping each other with very few gaps (e.g. between genes *C4B* and *C4A*). The array was accompanied by two custom-made files that were necessary for the analysis; the "gal" file and the "ids.txt" file.

Experiments were performed to confirm the efficiency of hybridisation throughout the array by Diego Ottaviani, Human Cytogenetics Laboratory.

2.18 DNA labelling, array hybridisation and washes

Similar quantities of test and control DNA within the range of 0.3-0.5 μg were labelled using the BioPrime Array CGH genomic labelling module (Invitrogen). Briefly, 60 μl of 2.5x Random primer mix was added to the appropriate amount of DNA and the volume was made up to 126 μl. The mixture was denatured at 95 °C for 10 minutes and snap-cooled on ice. 15 μl of 10x dCTP mix, 6 μl of Cy3 or Cy5 labelled dCTP (Perkin Elmer) and 3 μl Klenow fragment were added to the mixture while on ice, and the sample was incubated at 37 °C overnight. The unincorporated nucleotides were removed using MicroSpin™ G-50 Columns (Amersham). Fluorescent dye incorporation was monitored using the Nanodrop® ND-1000 spectrophotometer. In one tube, the two dye-labelled DNA samples were mixed together with 135 μl human Cot DNA (1 mg/ml, (Roche)) and precipitated with absolute ethanol following standard methodology. In a second tube, 80 μl of salmon or herring DNA (10 mg/ml) was mixed with 135 μl human Cot DNA and precipitated in a similar manner.

To resuspend the pellets in each of the two tubes, 400 μl of array Hybridisation buffer (2x SSC, 50% v/v deionised formamide, 10 mM Tris-HCl [pH 7.4], 10% w/v dextran sulphate, 0.1% v/v Tween 20) were used at 80 °C. After applying the buffer, the samples were kept at 80 °C on a heat block for 10 minutes with vortexing every 2 minutes. The contents of the second tube were then applied to the array which was

placed in a humidity chamber containing layers of absorbent paper soaked in a mixture of 20% v/v formamide and 2x SSC. The chamber was sealed air-tight and placed in a hybridisation oven with gentle rocking at 37 °C for 1 hour. Meanwhile, 40 ng of yeast tRNA (Sigma) were mixed with the contents of the first tube, which was also placed at 37 °C for 1 hour, while array pre-hybridisation was taking place. Subsequently, the contents of the first tube were applied to the array, which was sealed and allowed to hybridise for 48 hours at the same conditions as above.

The slide was removed from the hybridisation chamber and the following washes were performed in a coplin jar, on a shaking platform. 5 minutes with a solution of 2x SSC and 0.03% w/v SDS, at room temperature, followed by 5 minutes of the same solution at 65 °C. The slide was then rinsed briefly in 0.2x SSC at room temperature, followed by four 20-minute washes with the same solution, then a 10 minute wash in PBS with 0.05% v/v Tween 20, and a final brief rinse in distilled water. The slide was transferred to a suitable container which was lined up with absorbent paper at the bottom and was dried by centrifugation at 1000 rpm for 5 minutes in a swinging platform rotor.

2.19 Array scanning, data acquisition and analysis

Arrays were scanned with the help of Dr Ian Giddings, at the Cancer Research UK DNA microarray facility at the Institute of Cancer Research, in Sutton. A GenePix® 4000B microarray scanner (Molecular Devices) was used with accompanying GenePix Pro software. The dyes Cy3 and Cy5 were scanned at 532 nm and 635 nm respectively, and the voltage for each detector was adjusted in order to balance the signal from each channel. Two image files were generated, one for each dye, which were saved in tiff format.

The resulting image files were processed with the open access software, Spotfinder version 3.1.1. (available at: http://www.tm4.org/spotfinder.html). During this step a grid of squares was applied over each of the 24 sections of the array. Each square enclosed a spot, which represented a particular clone on the array. Squares which contained unspecific signal due to dirt etc, were manually excluded at this stage.

Values were taken of the average signal intensity within the spot (foreground) and the average signal intensity outside the spot (background). The values extracted were saved in spreadsheet format (.mev) and used for data analysis.

The analysis was performed with the help of Richard Mitter, at the Bioinformatics and Biostatistics group, London Research Institute Cancer Research UK, using an open-source package (Bioconductor - limma (Smyth 2005)). The data from the ".mev" files were read-in and mapped to relevant clone identifiers using the array-specific ".gal" file [Appendix C]. Each clone identifier was then mapped to specific annotation concerning its genomic location, held in the file "ids.txt" [Appendix D]. To obtain more accurate values of the hybridisation level for each spot, the background was subtracted from the foreground. To remove any systematic trends which arise from the microarray technology, rather than from differences between the DNA samples hybridized to the arrays, the data were normalised in two ways: within each array, and between different arrays.

The $\log_2$ratio values of the signal difference from the one dye relative to that of the other dye were calculated and normalised for dye-bias within each array ("Printip loess" method). This method also attempts to correct systematic variation, such as that induced by the use of multiple print-heads during array spotting. To calculate the combined mean for the mean $\log_2$ratio value of each MHC clone across different arrays, the inter-quartile range of those arrays was normalised to obtain a similar distribution of the $\log_2$ratio values using between-array normalisation ("Quantile" method).

Any spots not associated with a genomic location within the MHC, such as control clones and plant DNA, were removed from the analysis. The probes were grouped by clone identity and all non-numeric values, such as those cases where background was higher than foreground were removed, and the mean $\log_2$ratio was calculated. Normalised $\log_2$ratio values were produced in text form for positional visualisation in the UCSC genome browser.

Similar results were obtained in experiments where the dye-labelling of the test and control DNA samples was swapped.

To examine the relationship between MHC array data and genomic features such as: GC content, gene density, exon density, density of LINE or SINE elements, and CpG islands, sequence data for the region covered by the array were extracted from the human genome database at http://genome.ucsc.edu (hg17, May 2004 release). For gene and exon density the RefSeq genes schema was used, while the RepeatMasker schema was used to map LINE and SINE repeats.

Two different approaches were used for data smoothing. The "Running Mean" was calculated across a sliding window with sample size n=6 of the combined means of $\log_2$ratio values. To plot the "Loess" curve, a polynomial surface was fitted to the data using local fitting at 0.1 intervals.

Chapter 3

Replication timing in the human MHC

3.1 Introduction

A multitude of studies on replication timing have established the existence of various relationships: gene and GC-rich regions replicate early in S-phase while regions with few or no genes and high AT content replicate later. In addition, transcriptionally active genes generally replicate earlier than non-expressed genes.

These relationships have been reported over the years by studies employing different methodological approaches to investigate replication. Early experiments such as the replication banding of Chinese hamster chromosomes (Holmquist, Gray et al. 1982) paved the way for analysing replication in higher eukaryotic chromosomes. Replication bands were compared with cytogenetic bands, which reflect qualities such as GC richness and chromatin compaction of the different chromosomal regions. Centrifugal elutriation was applied as a means of separating sub-populations of cells with different DNA content to analyse their progression through S-phase (Dhar, Mager et al. 1988). Similarly, chemical synchronisation by block and release, and flow-sorting of cells into different S-phase sub-populations based on DNA content, were used in other studies, as described below. Finally, replication timing has been assessed by FISH, which enables calculation of the proportion of chromosomes in asynchronous S-phase cells that have replicated a particular genomic region (Selig, Okumura et al. 1992).

Over 30% of genes in the MHC encode immune-related proteins. The classical MHC class II genes are expressed constitutively in antigen presenting cells such as B-lymphocytes, macrophages and dendritic cells. Cytokines such as IFN-γ induce coordinated expression of these genes in non-expressing cells and upregulate several other genes across the MHC. In fact, up to 15 genes are induced or upregulated within ~600 Kb of the class II region.

Two studies of replication timing in the MHC have been reported previously. Tenzen *et al.* used synchronised myeloid leukaemia cells at the G1 to S-phase boundary with two successive treatments with aphidicolin (Tenzen, Yamagata et al. 1997). This drug inhibits the replicative DNA polymerase α in a reversible manner. Replicating cells are therefore blocked in S-phase whereas those in G2, M and G1 are allowed to cycle until they reach the G1/S border. Cells were labelled with radioactive nucleotides at specific times in order to label "old" and "new" DNA. At hourly intervals upon release from block the cells were also incubated with BrdU. "New" or nascent DNA was isolated according to radioactivity measurements and further purification by IP with anti-BrdU antibodies. The nascent DNA was then quantified by competitive PCR. A synthetic DNA molecule slightly longer in size than the normal PCR product was designed for each of their fifteen primer sets that were distributed across a 870 Kb region spanning the class II/III boundary. By adding equal amounts of "competitor" DNA in every PCR reaction – one for each of the six hourly intervals per primer pair – and an appropriate volume of nascent DNA based on equal scintillation counts, they were able to establish the replication timing for each locus.

In the Tenzen study, a primer set that spanned the class III gene *TNF* was found to replicate in the first S-phase interval (Figure 3.1). The two subsequent loci tested, overlapping ~150 Kb of the class III region, were shown to amplify most effectively in the second time interval. This was an indication that they mainly replicated during the second hour of S-phase, on average two hours earlier to the time the class II portion had been found to replicate. The authors suggest that the fourth time interval showed the strongest band in BrdU-enriched nascent DNA for the four consecutive primer sets that were located in the last 180 Kb under investigation, in the classical class II region. This implied that those loci predominantly replicated in the fourth hour of S-phase. The remaining eight primer sets tested were found to replicate in accordance with the transition from earlier to later replication, a phenomenon that the authors described as "precise switching of DNA replication timing".

In 2005, Woodfine and colleagues reported their findings on the replication timing of the MHC using a chromosome 6 tiling path array (Woodfine, Beare et al. 2005). In these arrays, which offered an average resolution of 79 Kb, the MHC was represented

by BAC clones derived from the HLA homozygous cell line, PGF. Nuclei of asynchronously cycling cells from a lymphoblastoid cell line were flow-sorted into G1 and S-phase according to DNA content, followed by DNA extraction and differential labelling with Cy3 and Cy5 fluorochromes. After suppression of repetitive sequences with Cot1 DNA, the two DNA fractions were co-hybridized onto the microarrays. Subsequently, the relative fluorescent intensities of the two dyes were quantified and analysed (Woodfine, Carter et al. 2005).

The principle of this assay is that the number of S-phase cells in which a particular sequence has replicated will be proportional to the time at which replication takes place. How early a particular sequence replicates, is thus reflected in the ratio of S-phase versus G1 DNA. According to the scaling criteria imposed on the array data, ratio values that approach 2 identified early replicating regions whereas those close to 1 indicated late replicating sequences.

The MHC was found to contain the earliest replicating DNA within chromosome 6 with a ratio of 1.95, while the latest replicating clone within this region had a ratio of 1.31. The early replicating pattern in the MHC class III region was found to be flanked by the later replicating pattern in the MHC class I and II regions (Figure 3.2). The authors did not describe sharp transitions in replication timing at the class II/III region boundary (Woodfine, Beare et al. 2005).

To explore the effect of transcription on replication timing in the MHC, a third study was performed in the Human Cytogenetics Laboratory (unpublished), using the FISH method first reported by Selig (Selig, Okumura et al. 1992). Forty two cosmid probes across the region were tested, with at least 500 nuclei scored for each probe. In S-phase nuclei, loci which have not yet replicated were visualised as single fluorescent signals or "singlets", and after replication as double fluorescent signals or "doublets", at each chromosome homologue. The proportion of singlets to doublets in a population of S-phase nuclei gave the average replication time for each locus tested.

Three different cell types were used in the study: i) the AHB B-lymphoblastoid cell line, with high levels of constitutive expression for the classical class II genes; ii) MRC5 fibroblasts that do not express any of the classical class II genes; and iii)

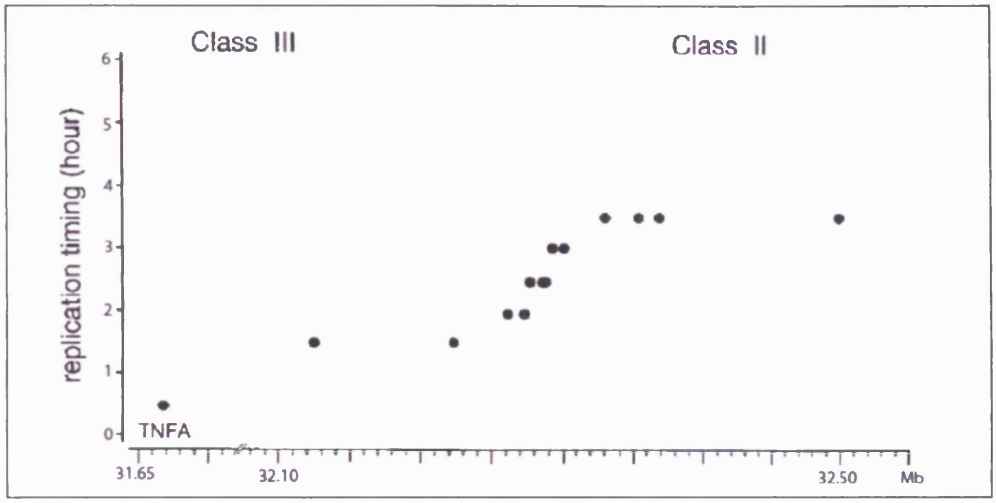


Figure 3.1 Replication timing at the MHC class II/III boundary (adapted from Tenzen et al. 1997).

A precise "switch" in replication timing was found at the MHC class II/III region boundary that apparently coincides with a transition in GC content.

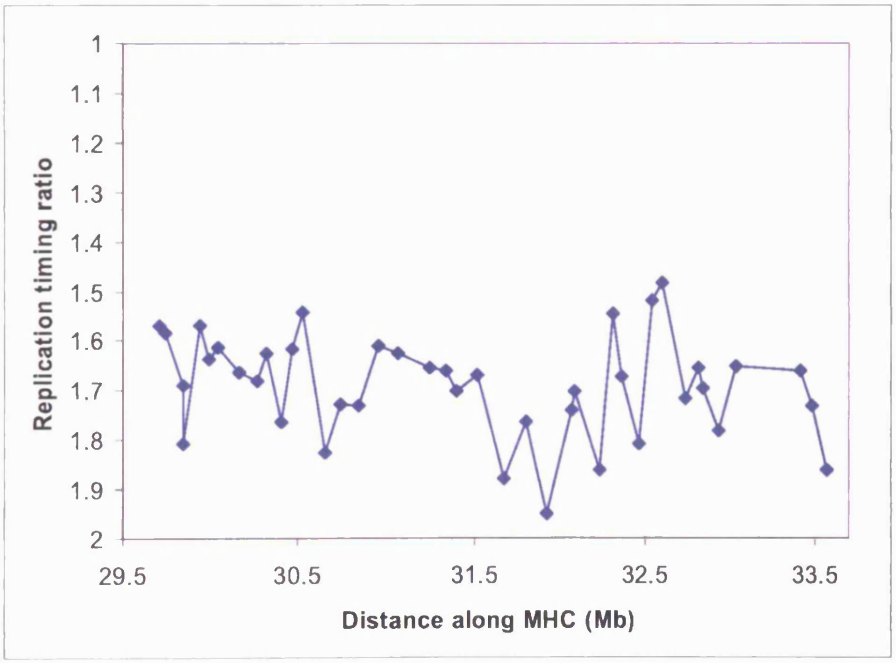


Figure 3.2 Replication timing in the MHC using a tile-path array of BAC clones (adapted from Woodfine et al. 2005).

The class III region contains the earliest replicating point (around 31.9 Mb) within the MHC, whereas the MHC class II region has the latest replicating point (around 32.6 Mb).

MRC5 fibroblasts that have been induced with IFN-γ to express classical class II genes.

Most of the ~3.6 Mb examined, using 42 probes, in AHB cells replicated early. However, the class II region between *HLA-DRA* (32.5 Mb) and the *TAP/PSMB* (32.8 Mb) cluster replicated later (Figure 3.3, blue line). Sharp transitions to earlier replication timing were seen on either side of the later replicating domains, while the earliest replication time was found to be in the class III region, close to the *HSPA1L* gene cluster.

FISH analysis of replication in MRC5 cells, using 30 probes, in the class II and III regions, before and after IFN-γ induction, showed a very similar pattern of replication timing to that of AHB cells. In MRC5 cells, DNA replication occurs slightly later than

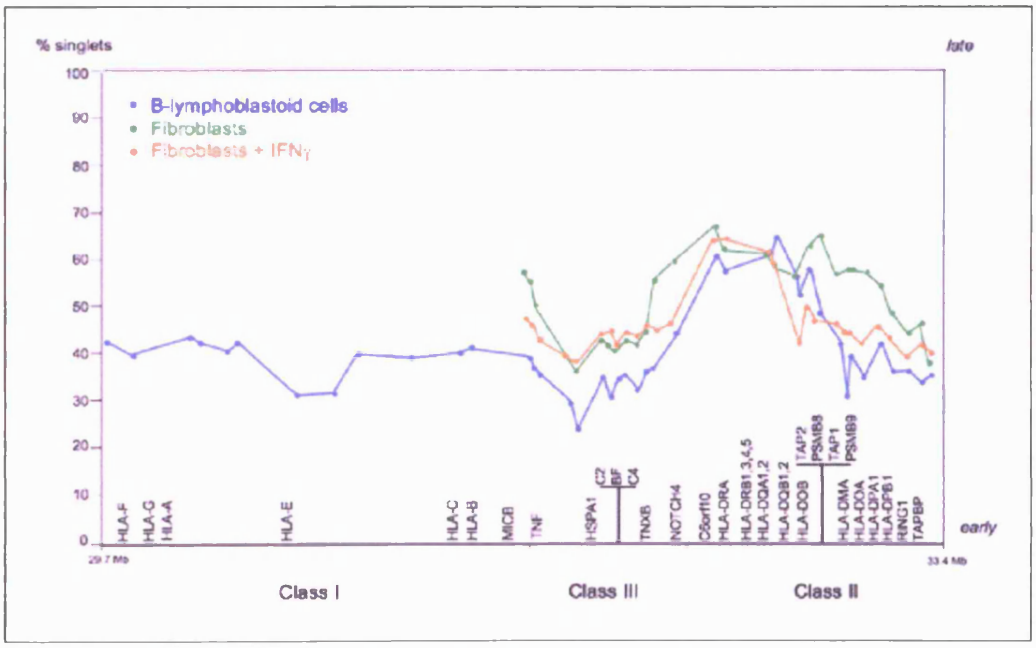


Figure 3.3 FISH analysis of replication timing in the MHC (data courtesy of P. Johonnett).

The data points show the relative replication time, expressed as percentages of singlet FISH signals, of probes tested in AHB cells (blue line), MRC5 cells (green line) and MRC5 cells induced with IFNγ (orange line). Additional information on the location of landmark genes and MHC class are also available on the horizontal axis.

in AHB cells for most of the class II and III regions (Figure 3.3, green line). In MRC5 cells that had been treated with IFN-γ for 24 h to induce expression of the classical class II genes, there was a small shift to earlier replication around the class II/III boundary and in the class II region from *HLA-DQB* to *TAPBP* (Figure 3.3, orange line). Fibroblasts before and after induction showed a remarkably similar replication time in the class II region from *HLA-DRA* to *HLA-DQA,* and in most of the class III region.

In spite of the active expression status of the classical class II genes in the B-cells and the IFN-γ induced fibroblasts, the replication timing for this genomic region did not conform to the concept that transcriptionally active sequences replicate early. Furthermore, the suitability of FISH as a method for replication timing analysis has received criticism because of its resolution. Does the region encompassing the classical MHC class II genes really replicate late regardless of expression status, or was this finding a methodological artefact? An alternative assay was thus developed and the relationship between replication timing and gene activity examined at ten different loci across the MHC class II and III regions. The principle of the assay used was to assess nascent DNA enrichment in cells flow-sorted into different S-phase sub-populations according to DNA content.

3.2 Results

3.2.1 Analysis of replication timing in B-lymphoblastoid cells

Firstly, to assess the progress of replication in B-lymphoblastoid cells, an initial investigation was carried out on the replication pattern of AHB cells that had been sorted with respect to DNA content (Figure 3.4). Coverslip fixed nuclei for each of the four S-phase fractions and G1, were examined using confocal microscopy. S1 nuclei had several small foci distributed throughout the nucleoplasm (Figure 3.5). Progressively in S2 and S3 foci became larger in size and were more abundant in the nuclear periphery. Finally, in S4 replication was confined to a few large replication foci. This nuclear pattern of replication foci is similar to that reported previously

(Nakamura, Morita et al. 1986; Nakayasu and Berezney 1989; van Dierendonck, Keyzer et al. 1989; O'Keefe, Henderson et al. 1992).

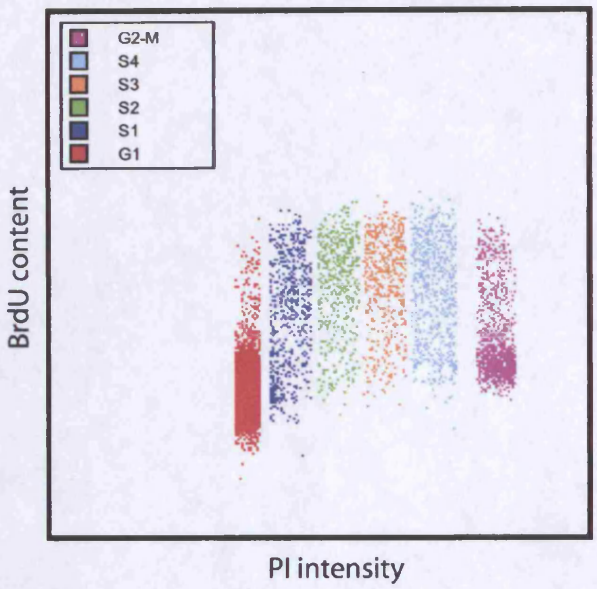


Figure 3.4 Typical FACS profile of PI intensity, corresponding to DNA content, with respect to BrdU content for an AHB cell population.

G1, G2/M and four sub-populations of S-phase cells, S1-S4, were flow-sorted according to DNA content (refer to Figure 3.6). Using the same sorting criteria cells were also analysed for BrdU content, to enable subsequent identification of replicating cells by confocal microscopy. It is noteworthy that both the G1 and G2/M fractions contain cells that are BrdU positive.

To optimise the replication timing assay for this B-lymphoblastoid cell line, modifications were made to a published protocol (Hansen, Canfield et al. 1993). As it was important to achieve a good signal to noise ratio it was necessary to establish, firstly, a suitable time span for BrdU incorporation, and secondly, the number of cells that had to be sorted in each sub-population to achieve consistent results. Briefly, an asynchronous population of AHB cells was incubated in the presence of BrdU for 1 hour after which the cells were fixed. Following fixation the sample was first treated with RNase and then the DNA stained with PI (Krishan 1975). Subsequently, equal numbers of cells (10^5) were sorted in each of six cell cycle sub-populations G1, G2/M and four fractions of S-phase (S1-S4), according to DNA content (Figure 3.6).

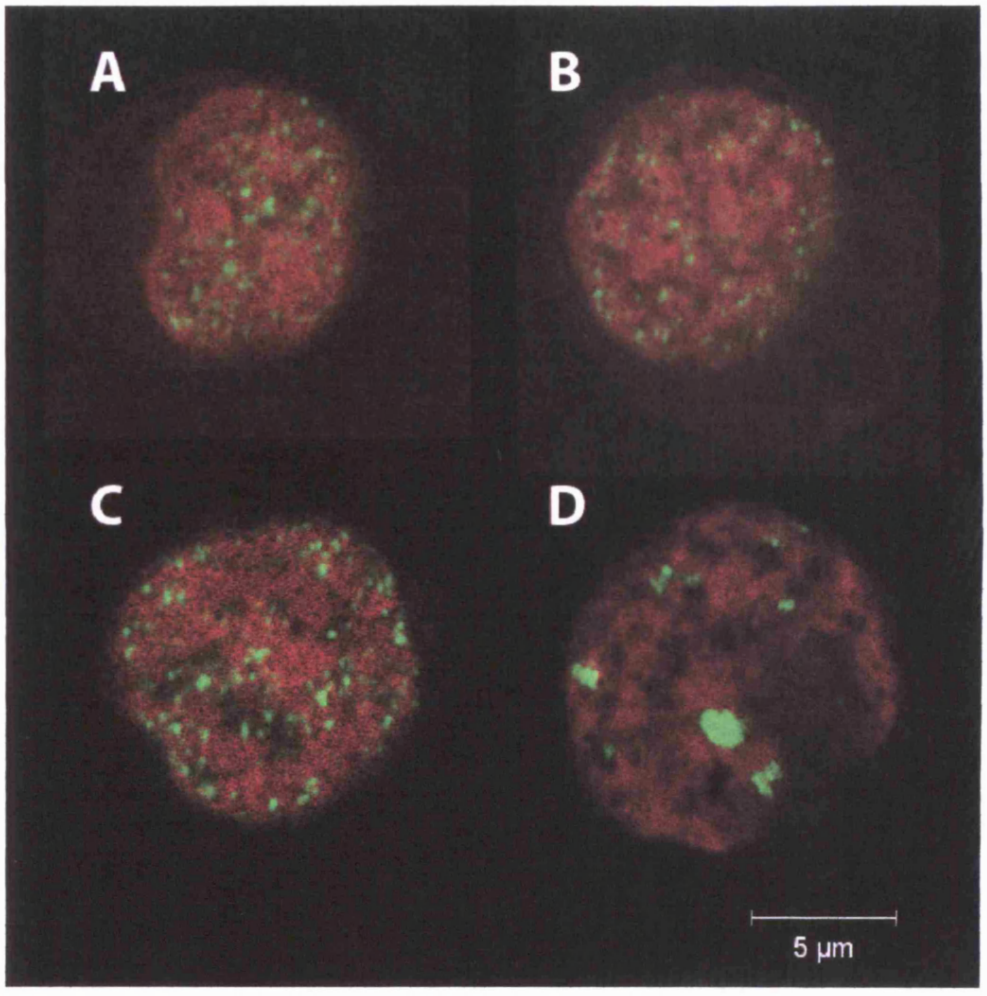


Figure 3.5 Replication patterns in nuclei of AHB cells.

Different patterns of DNA replication were identified in AHB cells sorted in four fractions of S-phase according to DNA content, followed by immunofluorescent detection of BrdU using confocal microscopy. Panels A-D, representative of cells in S1-S4, progress from patterns with small foci within the nuclear core, to more enlarged foci at the nuclear periphery, and finally very few large regions of replication.

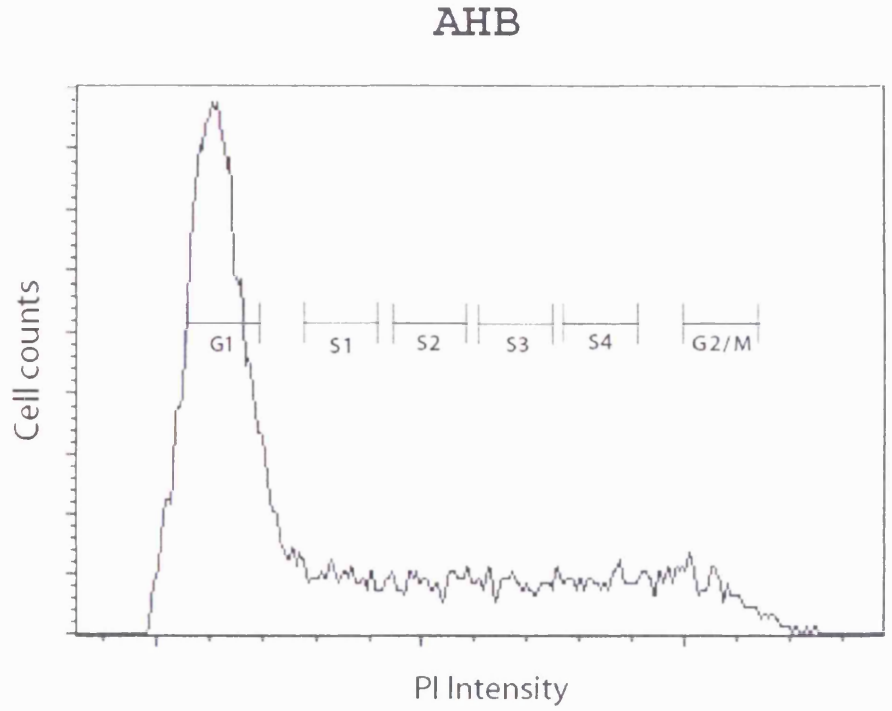


Figure 3.6 Typical FACS profile of PI intensity, corresponding to DNA content, for AHB cells.

G1, G2/M and four sub-populations of S-phase cells, S1-S4, were flow-sorted according to DNA content. Each cell cycle fraction was demarcated by a pair of vertical lines representing the flow-sorting gates used for this and similar experiments. The S1-S4 gates have identical lengths and therefore divide S-phase to equal parts. Gaps were left between the gates to increase the purity of each fraction.

To ensure uniform recovery of BrdU-enriched DNA across the six different fractions following IP, PCR examination of ~200 bp of mitochondrial (mt) DNA was used (Gomez and Brockdorff 2004). Mitochondrial DNA replication has been found to continue throughout the cell cycle in mammalian cells even though, surprisingly, the number of mitochondria in each cell remains roughly the same (Bogenhagen and Clayton 1977; Davis and Clayton 1996). Bands with similar intensity for each of the six fractions were seen, illustrating that the IP had worked with similar efficiency (Figure 3.7).

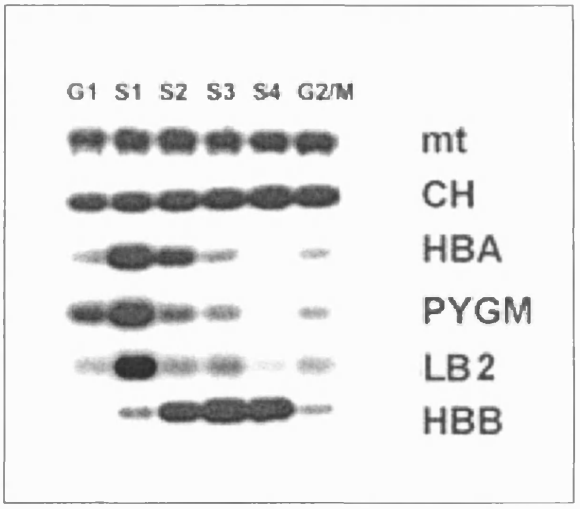


Figure 3.7 Relative abundance of nascent DNA at control loci in cell-cycle fractions of AHB cells.

Mitochondrial DNA abundance was examined as a measure of efficiency for immunoprecipitated BrdU-enriched DNA. Similar band intensities across the six fractions indicate that the IP procedure has worked with similar efficiency. High levels of nascent DNA enrichment were observed in the S1 fraction of the cell-cycle for the early-replicating *PYGM*, *HBA* and *LB2*, while *HBB* was enriched in later fractions of S-phase. Chinese hamster DNA was used to "spike" the different S-phase fractions as an additional method of monitoring the efficiency of BrdU IP. All fractions were enriched equally at the end of successful IP experiments. Results shown are typical of at least two independent experiments.

PCR amplification of regions of the genome that had been reported previously to replicate early or late were used to ensure that the six fractions collected were indeed representative of early or late replicating DNA sequences. These were at or near the genes: glycogen phosphorylase (*PYGM*), α globin (*HBA*), lamin B2 (*LB2*) – early – and β globin (*HBB*) – late – controls (Selig, Okumura et al. 1992; Giacca, Zentilin et al. 1994; Smith and Higgs 1999) [Appendix E].

As expected, clear enrichment at early S-phase fractions was shown for *PYGM*, *HBA* and *LB2*. *HBB* on the other hand, was enriched in late S-phase fractions (Figure 3.7).

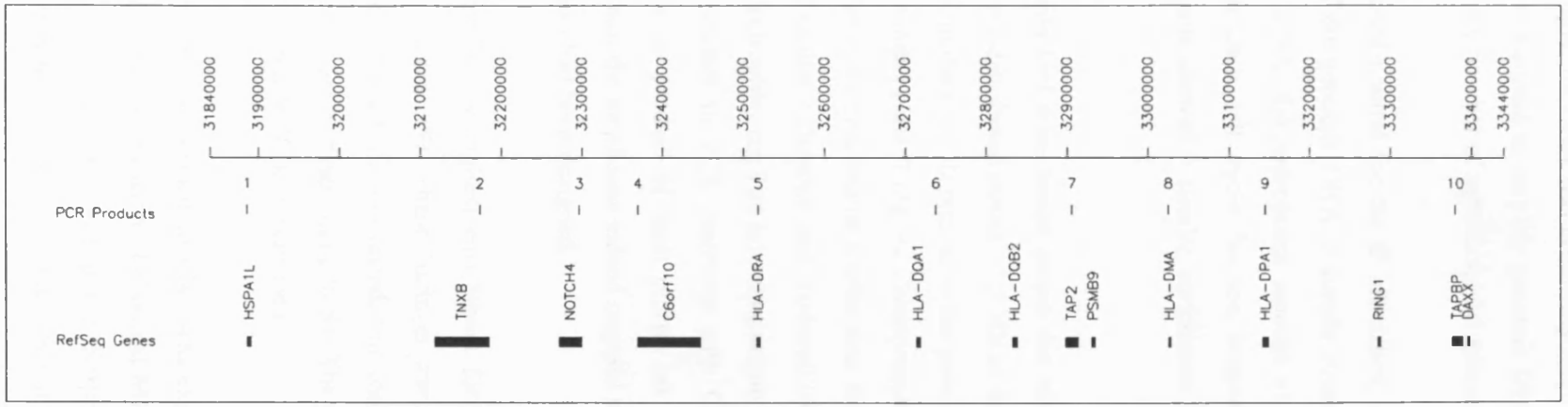


Figure 3.8 Positions of primer sets designed across ~1.5 Mb of the human MHC.

Above, the scale indicates the base position for this part of chromosome 6. The PCR products of the primer sets designed are represented with numbers 1-10 followed by short vertical lines. Below, based on the schema of RefSeq genes at the http://genome.ucsc.edu/ human genome database, landmark genes in the vicinity of the products are represented as black boxes. Subsequently, all primer sets / PCR products are identified by the name of the gene nearest to them.

Interestingly, mt DNA was efficiently amplified by 19 cycles, while the primer pairs that were designed to amplify genomic DNA by 26 cycles. This is likely due to the higher copy number of mitochondrial genome in each cell.

As a second control for the IP efficiency, using a number of PCR cycles similar to those of the genomic DNA, a simple strategy was adopted. BrdU-enriched Chinese hamster DNA, the appropriate amount of which was determined empirically, was added to each cell cycle fraction immediately after flow-sorting. Successful IP experiments showed a similar enrichment in CH DNA for all six fractions (Figure 3.7).

Ten primer loci were tested across the MHC class II and III regions (Figure 3.8). These were distributed across ~1.5 Mb of the MHC, from the heat shock protein gene, *HSPA1L*, in the class III region, to the gene encoding tapasin binding protein, *TAPBP*, in the extended class II region. Considerable care was taken to ensure that the primer pairs produced amplicons of similar size. Similarly, the PCR reaction for each primer pair had similar efficiency and produced bands with similar intensity when analysed by gel electrophoresis. The primers designed to amplify human genomic DNA did not give a product in PCR reactions with Chinese hamster DNA. Furthermore, the amplification product of each primer set for the MHC region was sequenced to confirm that the amplicons indeed mapped to the specified genomic regions for which the primers had been designed.

Unexpectedly, the highest enrichment for all the MHC loci tested was in the S1 fraction (Figure 3.9). These findings were consistent in at least two independent experiments for all the loci tested. For some loci, less enriched bands were present adjacent to those corresponding to S1. The pattern of the faint bands observed varied between the ten MHC loci examined.

These findings conformed to what was expected; a B-lymphocyte derived cell line with constitutive expression of classical MHC class II genes had been anticipated to replicate those loci early in S-phase. Surprisingly though, these results were not in agreement with the findings of the FISH study which had found the classical class II region to replicate late in S-phase for AHB cells. According to the FISH study the

sequences adjacent to the classical class II region replicated early in S-phase. This finding is in agreement with the results presented here. What are the underlying reasons for the observed difference between the two studies? Would similar findings recur for a cell type with a different expression profile across the MHC?

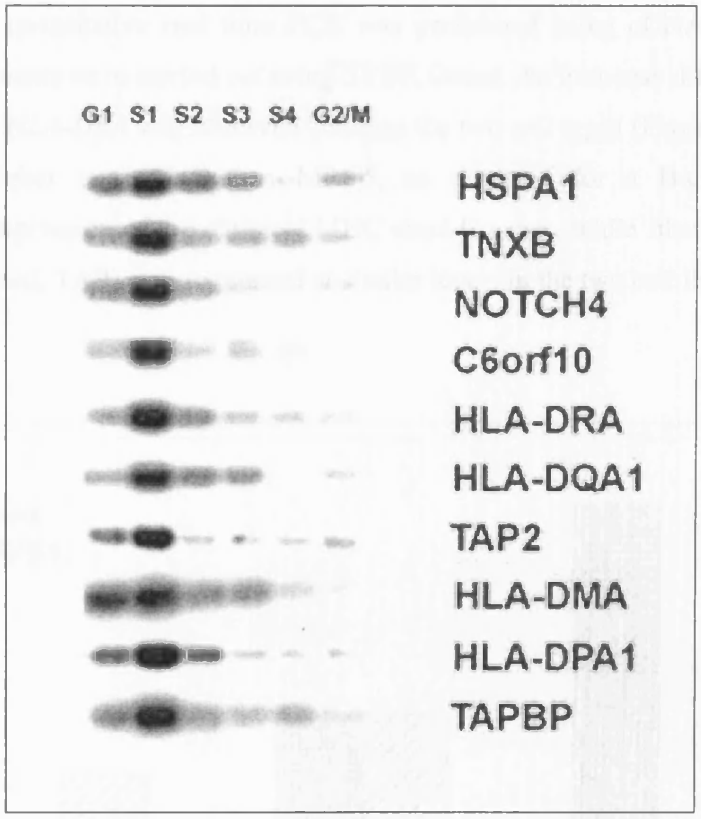


Figure 3.9 Relative abundance of nascent DNA at ten MHC loci in cell-cycle fractions of AHB cells.

The MHC loci examined were most enriched in the S1 fraction, indicating that the MHC replicates in the first half of S-phase. Results shown are typical of at least two independent experiments.

3.2.2 Replication timing in fibroblasts

The same experimental approach described for AHB cells above was used to examine replication in fetal lung fibroblasts, which had previously been used in the FISH study. Potential differences in the findings of the two approaches could thus be

assessed. Would the silence of the classical MHC class II genes affect replication timing in these cells?

The expression of *HLA-DRA*, a gene representative of the classical MHC class II region, was used to examine expression differences between fibroblasts and lymphoblastoid cells. RNA was extracted from both cell lines and following reverse transcription, quantitative real time PCR was performed using cDNA as template. These experiments were carried out using SYBR Green. An immense difference in the expression of *HLA-DRA* was observed between the two cell types (Figure 3.10). AHB has much higher expression than MRC5, as expected for a B-cell line with constitutive expression of the classical MHC class II genes, while fibroblasts do not. Also as expected, TAP1 was expressed at similar levels in the two cell lines.

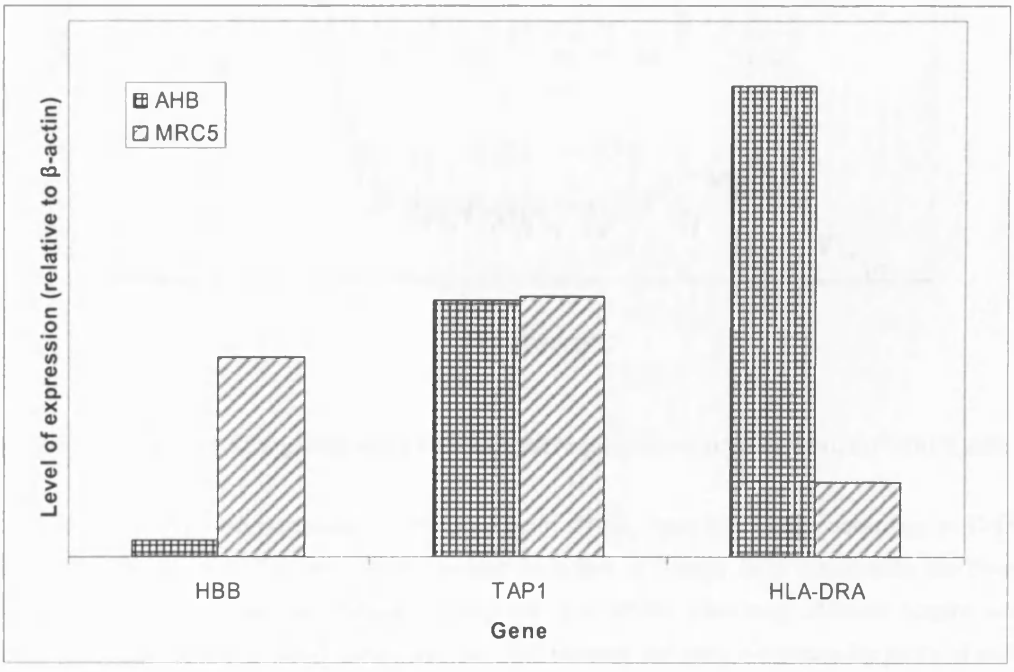


Figure 3.10 Relative expression of genes HLA-DRA, TAP1 and HBB in fibroblasts and B-lymphoblastoid cells.

The *HLA-DRA* transcript was considerably more abundant in AHB cells than MRC5 cells. On the other hand, MRC5 cells had higher levels of *HBB* expression than the AHB cell line, while *TAP1* was found to be expressed at similar levels.

A typical FACS profile of MRC5 cells that was used to sort cell-cycle fractions is shown in Figure 3.11. Noticeably, these adherent cells had a slower doubling rate than AHB cells and were not as easy to handle. In particular, they presented problems in the flow sorting procedure as cells often clumped together, blocking the nozzle of the FACS machine and delaying the sorts.

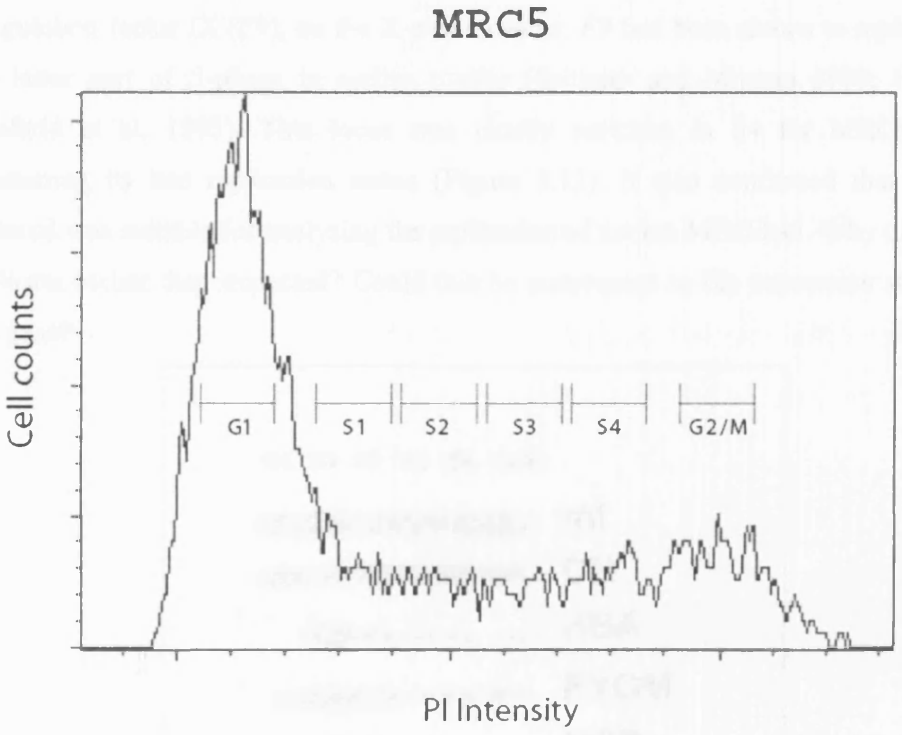


Figure 3.11 Typical FACS profile of PI intensity, corresponding to DNA content, for MRC5 cells.

G1, G2/M and four sub-populations of S-phase cells, S1-S4, were flow-sorted according to DNA content. Each cell cycle fraction was demarcated by a pair of vertical lines representing the flow-sorting gates used for this and similar experiments. The S1-S4 gates have identical lengths and therefore divide S-phase to equal parts. Gaps were left between the gates to increase the purity of each fraction.

Replication timing of the control loci was examined first. *HBA* and *PYGM* were predominantly enriched in the S1 fraction of S-phase (Figure 3.12). Unexpectedly, however, the late replicating control *HBB* was mostly enriched in the S2 fraction. This result contradicted previous findings on the replication of this locus, and raised

questions regarding the precision of the flow sorting method. A possible explanation for this observation was that the particular locus in MRC5 cells indeed replicated earlier than expected and the sub-populations sorted were in fact accurate. This assumption was evaluated with two different approaches.

Firstly, various loci that could be used as alternative controls for late replication were assessed. Primers were designed to span a section of the ~33 Kb gene encoding coagulation factor IX (F9), on the X-chromosome. *F9* had been shown to replicate in the latter part of S-phase in earlier studies (Schmidt and Migeon 1990; Hansen, Canfield et al. 1993). This locus was mostly enriched in S4 for MRC5 cells, illustrating its late replication status (Figure 3.12). It also confirmed that the IP material was suitable for analysing the replication of the ten MHC loci. Why did *HBB* replicate earlier than expected? Could this be consequent to the expression status of the gene?

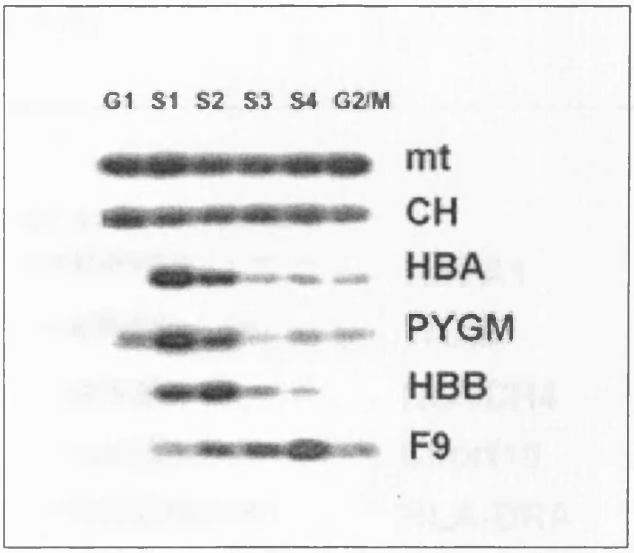


Figure 3.12 Relative abundance of nascent DNA at control loci in cell-cycle fractions of MRC5 cells.

Mitochondrial DNA abundance was examined as a measure of efficiency for immunoprecipitated BrdU-enriched DNA. Similar band intensities across the six fractions indicate that the IP procedure has worked with similar efficiency. Chinese hamster DNA was used to "spike" the different S-phase fractions as an additional method of monitoring the efficiency of BrdU IP. All fractions were enriched equally at the end of successful IP experiments. High levels of nascent DNA enrichment were observed in the early S-phase fractions for the early-replicating *PYGM* and *HBA*. Surprisingly, *HBB* was also enriched in early S-phase fractions. *F9*, which was used as an alternative late replicating control, was mostly enriched in the S4 fraction. Results shown are typical of at least two independent experiments.

As a second approach, *HBB* gene expression was examined in MRC5 and AHB cells by quantitative real time PCR. Considerably higher levels of expression for this gene were found in MRC5 compared to AHB cells (Figure 3.10). Evidence for *HBB* expression in fetal lung tissue, based on a microarray study, was found at: http://genome.ucsc.edu/cgi-bin/hgGene?hgsid=75139125. Given that MRC5 cells are derived from fetal lung, these data support the PCR findings of this study.

Overall, these experiments showed that the BrdU-IP with flow-sorted sub-populations of MRC5 cells was suitable for further analysis.

Examination of the ten MHC loci using BrdU-enriched DNA from MRC5 cells gave different results to those of the AHB cell line. Here, the ten loci were enriched in both the S1 and S2 sub-populations, while the *HLA-DRA* locus had a prominent additional band in S3 (Figure 3.13).

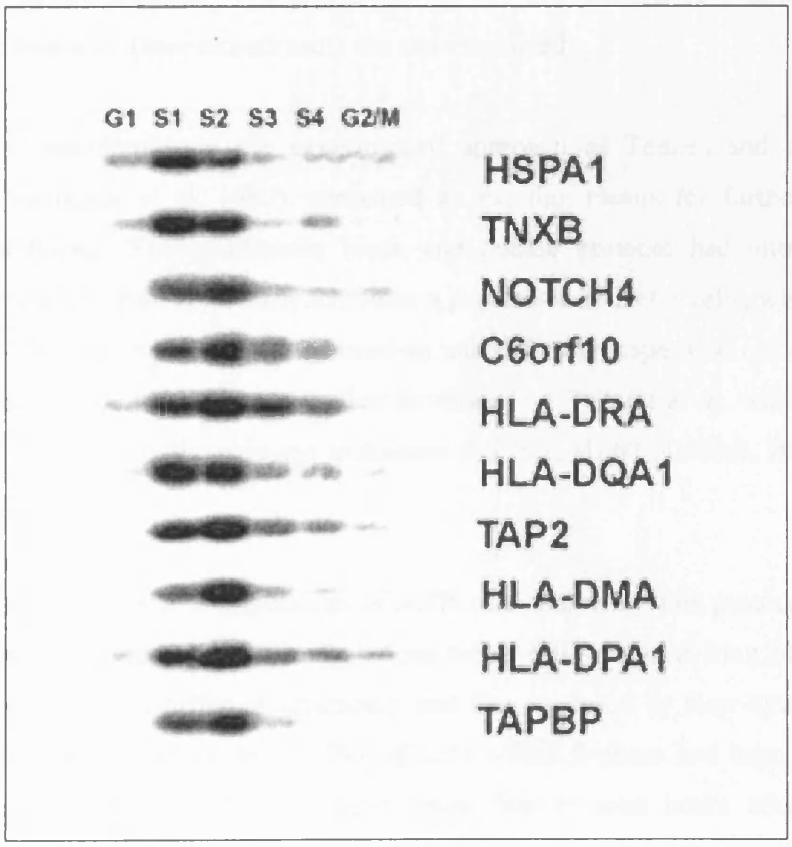


Figure 3.13 Relative abundance of nascent DNA at ten MHC loci in cell-cycle fractions of MRC5 cells.

The MHC loci examined were most enriched in the S1 and S2 fractions, indicating that the MHC replicates in the first half of S-phase. Results shown are typical of at least two independent experiments.

Was the MHC replication generally delayed in MRC5 relative to AHB cells due to lack of expression of the classical class II genes? The underlying reason for the difference in the timing of replication of the ten MHC loci between MRC5 and AHB cells is not entirely clear. In disagreement with the FISH study, these data strongly suggest that the MHC loci tested replicate in the first half of S-phase.

3.2.3 Replication analysis in a myeloid leukaemia cell line

This section describes experiments that were performed to analyse replication timing using cell synchronisation. The arguments for the introduction of a new cell line during the course of these experiments are also explained.

The use of aphidicolin in the experimental approach of Tenzen and colleagues (Tenzen, Yamagata et al. 1997), presented an exciting means for further refining replication timing. The aphidicolin block and release protocol had initially been applied by Pedrali-Noy *et al.* to synchronise a population of HeLa cells, where it was reported to be simple and able to be used on adherent and suspension cells (Pedrali-Noy, Spadari et al. 1980). It was further developed by Tribioli *et al.* who used this method to synchronise the myeloid leukaemia cell line, HL60 (Tribioli, Biamonti et al. 1987).

Attempts to synchronise a population of AHB cells following this protocol did not achieve the objective. This became obvious when cells released from block were collected and fixed at different timepoints and then analysed by flow cytometry. A considerable enrichment in the number of cells within S-phase had been expected, particularly for samples that had been fixed five to nine hours after release. Surprisingly, flow cytometric analysis revealed a similar pattern for samples within

that time frame. This was comparable to an asynchronous population of AHB cells, leading to the assumption that AHB cells were not amenable to aphidicolin-induced replication block and therefore synchronisation.

HL60 cells were then selected for analysis as this cell line had successfully been synchronised in the past (Tribioli, Biamonti et al. 1987; Tenzen, Yamagata et al. 1997). Use of MRC5 cells was avoided, as these adherent cells grow at a much slower rate and cannot be conveniently manipulated. To evaluate the suitability of this cell line for inclusion in this study, it was necessary: i) to investigate the expression status for classical MHC class II genes in HL60 cells, and ii) to compare replication timing in the MHC between the AHB, MRC5 and HL60 cells.

To assess the expression of classical MHC class II genes in HL60 cells, immunocytometric analysis was used to examine the presence of HLA-DRA antigen on the outer cell membrane. It was found that *HLA-DRA* is not expressed in this myeloid leukaemia cell line (Figure 3.14). As the class II genes are coordinately regulated, these data suggest that all the classical class II genes are silent in HL60.

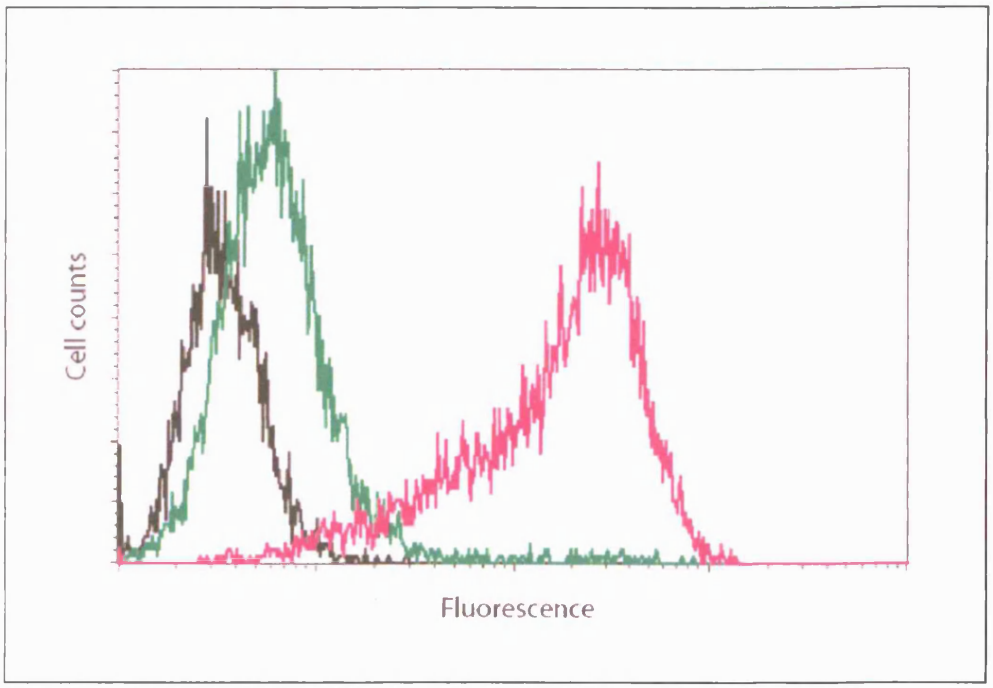


Figure 3.14 Immunofluorescence analysis of *HLA-DRA* expression in HL60 cells.

Black: HL60 cells without anti-DRA antibody (negative control); Green: HL60 cells incubated with FITC conjugated anti-DRA antibody show insignificant increase in amount of fluorescence; Red: AHB cells incubated with FITC conjugated anti-DRA antibody are shifted along the fluorescence axis (positive control).

An identical method to that used to examine replication timing for AHB and MRC5 cells was employed for the HL60 cells to establish a point of reference for the replication behaviour of HL60 cells with respect to AHB and MRC5 cells. An asynchronous population of HL60 cells was sorted in six sub-populations (Figure 3.15) followed by IP of BrdU-enriched DNA and replication timing analysis at the ten MHC loci.

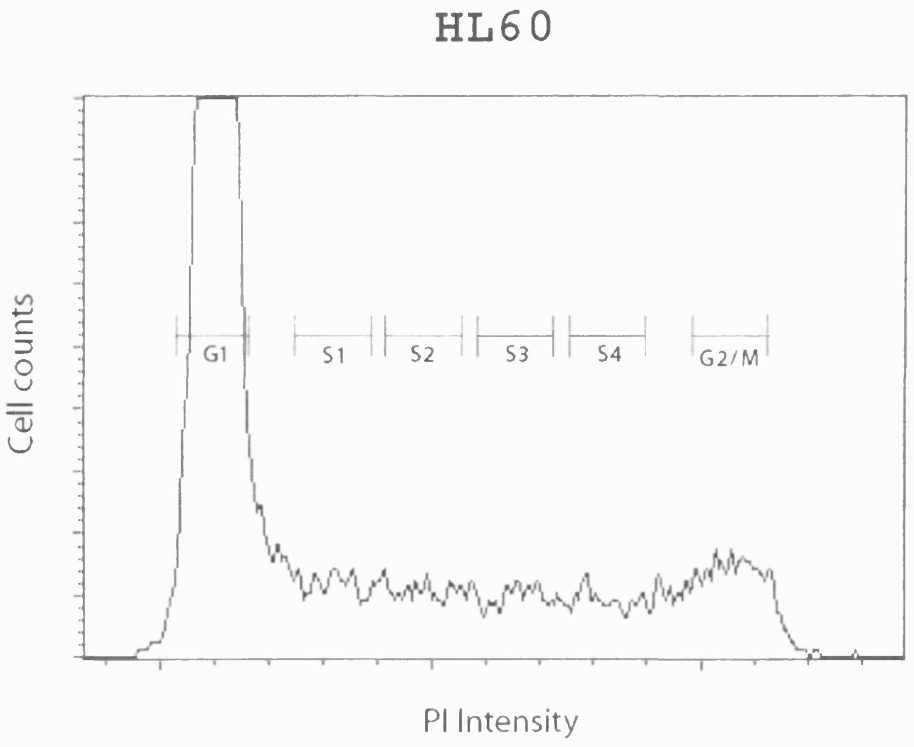


Figure 3.15 Typical FACS profile of PI intensity, corresponding to DNA content, for HL60 cells.

G1, G2/M and four sub-populations of S-phase cells, S1-S4, were flow-sorted according to DNA content. Each cell cycle fraction was demarcated by a pair of vertical lines representing the flow-sorting gates used for this and similar experiments. The S1-S4 gates have identical lengths and

therefore divide S-phase to equal parts. Gaps were left between the gates to increase the purity of each fraction. Results shown are typical of at least two independent experiments.

Examination of the control loci gave the expected results indicating that the flow-sorting procedure and IP protocol had worked optimally. The mt DNA and CH primers had signals of equivalent intensity across the six fractions. *HBA* and *PYGM* had strong enrichment in S1, whereas the *HBB* locus was mostly enriched in the S4 fraction (Figure 3.16).

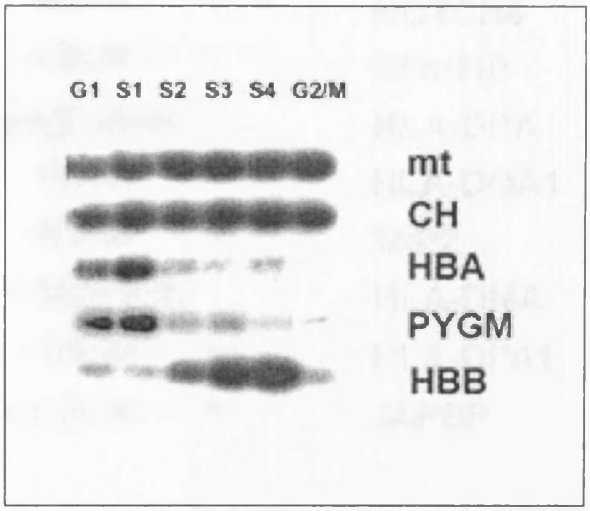


Figure 3.16 Relative abundance of nascent DNA at control loci in cell-cycle fractions of HL60 cells.

Mitochondrial DNA abundance was examined as a measure of efficiency for immunoprecipitated BrdU-enriched DNA. Similar band intensities across the six fractions indicate that the IP procedure has worked with similar efficiency. Chinese hamster DNA was used to "spike" the different S-phase fractions as an additional method of monitoring the efficiency of BrdU IP. All fractions were enriched equally at the end of successful IP experiments. High levels of nascent DNA enrichment were observed in the G1 and S1 fractions of the cell-cycle for the early-replicating *PYGM* and *HBA*, while *HBB* was enriched in the later fractions of S-phase, at S3 and S4. Results shown are typical of at least two independent experiments.

The ten MHC loci tested were found to replicate in the first half of S-phase in HL60 cells. All the loci had a prominent band in S1 and some had a second band in S2. For some loci the band in S2 was almost as strong as the S1 band (Figure 3.17). This pattern of enrichment resembled that of the MRC5 cells more than AHB. Taken together, these findings and those described previously for the AHB and MRC5 cells

suggest that the replication timing for the MHC class II and III regions is broadly similar in the three cell types.

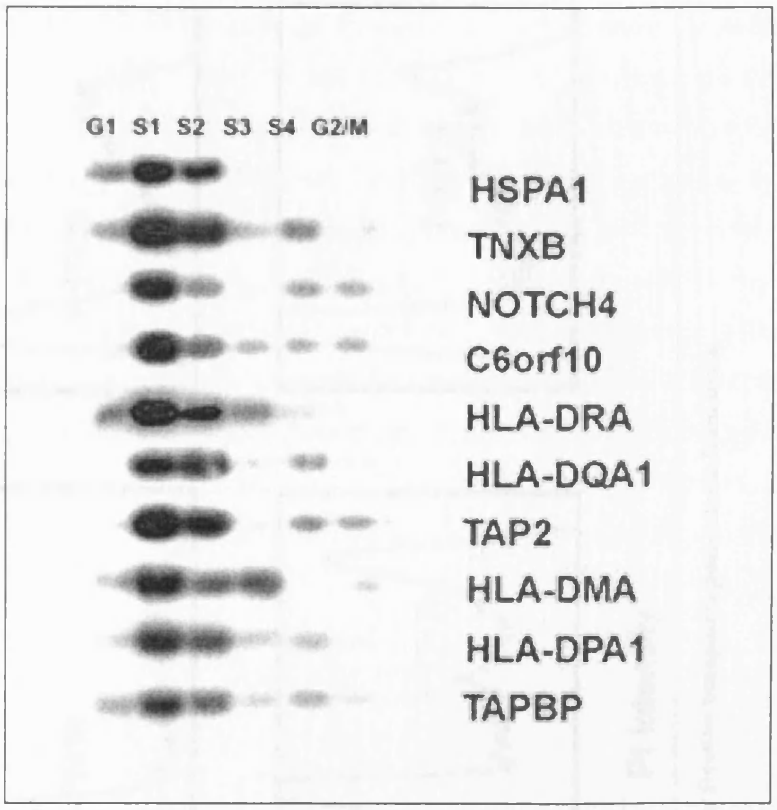


Figure 3.17 Relative abundance of nascent DNA at ten MHC loci in cell-cycle fractions of HL60 cells.

The MHC loci examined were most enriched in the S1 and S2 fractions, indicating that the MHC replicates in the first half of S-phase. Results shown are typical of at least two independent experiments.

In view of the similarities described between AHB, MRC5 and HL60 cells for the replication timing of the ten MHC loci, an alternative approach was taken to give more precise information on HL60. It was hoped that these experiments would also offer insights into the replication timing of AHB and MRC5 cells. To synchronise HL60 cells at the G1/S-phase boundary, an asynchronous population was incubated with aphidicolin. The drug was then removed and the cells resuspended and incubated in fresh media, prior to a second aphidicolin treatment. The drug was removed again and the cells were once more resuspended in fresh media. This was considered to be

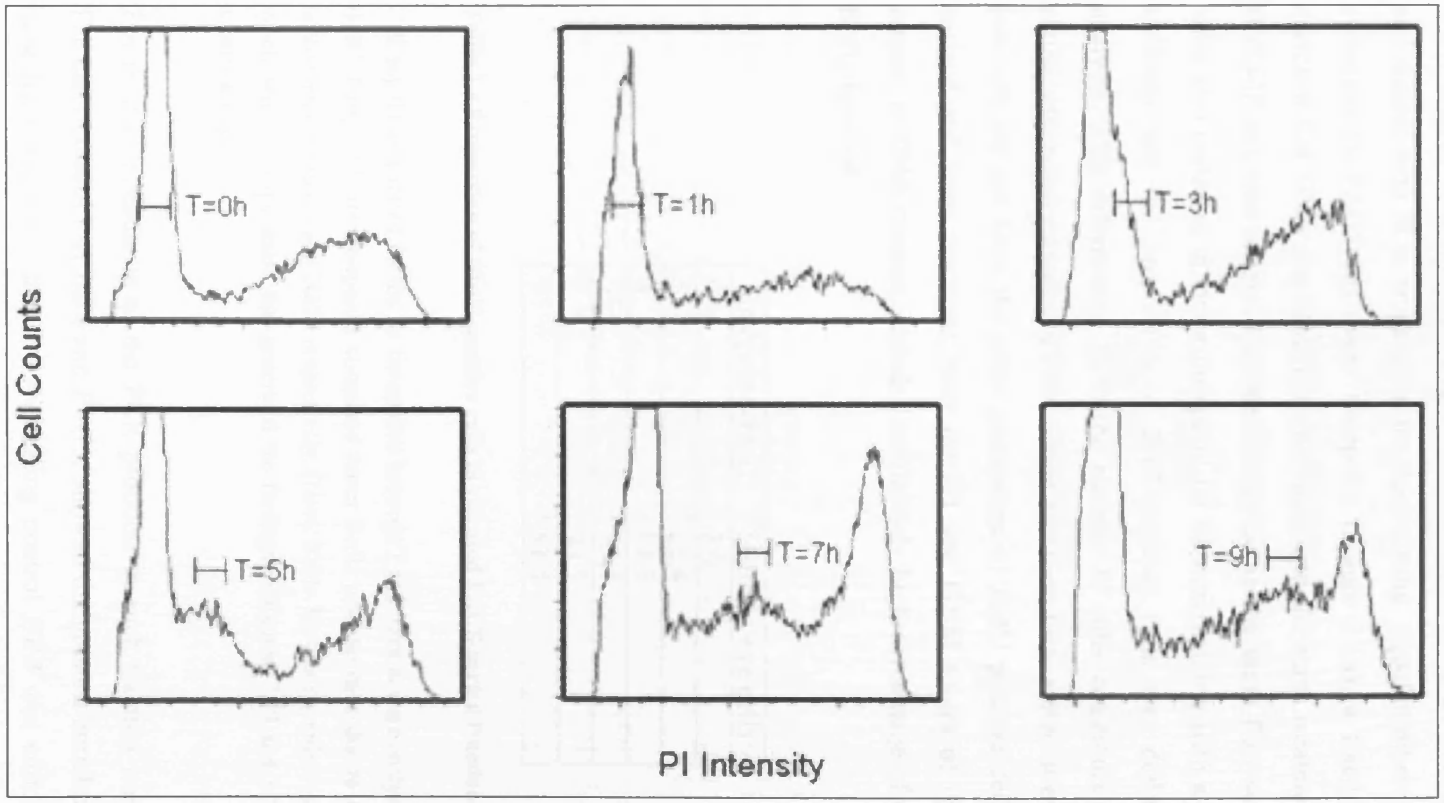


Figure 3.18 Typical FACS profiles of synchronised HL60 cells at different fixation timepoints post-release from block.

Cell counts are shown across varying PI intensities, corresponding to DNA content. Cells were labelled with BrdU for 1 hour prior to fixation at timepoints T=0, 1, 3, 5, 7 and 9 hours, after release from block. At each timepoint, identical numbers of cells were collected by flow-sorting using the appropriate gates around the peaks of synchronized cells demarcated by the pair of vertical lines on each plot.

timepoint t=0 hours. Some cells were fixed at this point, and then at timepoints t=1, 3, 5, 7 and 9 hours. Prior to fixation at the various timepoints, the respective samples were labelled with BrdU for 1 hour. After fixation, samples were treated with RNase and stained with PI in preparation for flow-sorting. Equal numbers of cells (10^5) were collected by FACS for every timepoint (Figure 3.18), a fixed volume of BrdU-enriched CH DNA was added in each one, and the same treatment, with respect to BrdU-IP etc, was followed as previously. Using the same flow-sorting criteria, cells were also collected at every timepoint for immunostaining with a suitable anti-BrdU antibody and the proportion of BrdU-positive cells was defined by cytometric analysis. This information gave the number of cells contributing to the observed signal, given that samples of cells which had been flow-sorted at each of the different intervals did not have the same percentage of BrdU positive cells (Table 3.1) The earliest and latest fractions, near the G1 and G2/M stages of the cell cycle, with respect to DNA content, included a relatively high percentage of cells that were not BrdU-labelled.

Timepoint (hrs)	% BrdU +ve cells
0	3.6
1	22.4
3	84.5
5	95.4
7	90.5
9	81.4

Table 3.1 Proportion of BrdU positive cells in released FACS-sorted fractions.

Cell populations sorted for the six timepoints between 1 and 9 hours were analyzed for the presence of BrdU. Early and late timepoints contained fewer BrdU positive cells due to their proximity to the unlabelled cells in G1 and G2/M respectively. (Note: Table 3.1 was compiled with data from the same synchronisation experiment that generated the findings in figures. 3.19 and 3.20, which typify these experiments).

The relative abundance of the PCR products in each fraction was then determined. The early control loci *HBA* and *PYGM* showed the greatest enrichment in the 1 and 3 hour fractions, while the late replicating control *HBB* was enriched in the 9 hour fraction (Figure 3.19).

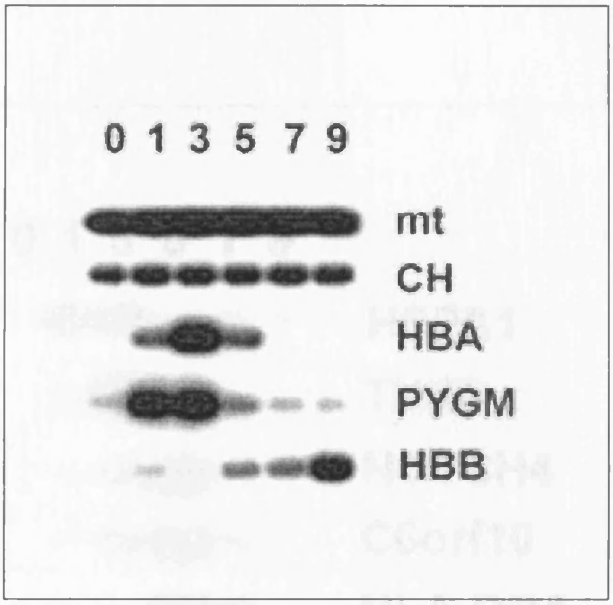


Figure 3.19 Relative abundance of nascent DNA at control loci for different fixation timepoints post-release from block in HL60 cells.

Mitochondrial DNA abundance was examined as a measure of efficiency for immunoprecipitated BrdU-enriched DNA. Similar band intensities across the six fractions indicate that the IP procedure has worked with similar efficiency. Chinese hamster DNA was used to "spike" the different S-phase fractions as an additional method of monitoring the efficiency of BrdU IP. All fractions were enriched equally at the end of successful IP experiments. High levels of nascent DNA enrichment were observed in the T=1 and T=3 timepoints for the early-replicating *PYGM* and *HBA* loci, while *HBB*, which replicates late, was mostly enriched at T=9. Results shown are typical of at least two independent experiments

The loci from *HSPA1* in the class III region to *HLA-DRA* in the classical class II region were found to replicate progressively later, while those from *HLA-DQA* to *TAPBP* in the class II region replicate earlier (Figure 3.20). The MHC class II region had a zone of later replication, in agreement with the FISH and other studies. This, however, was still within the first half of S-phase as the most enriched band of the latest replicating locus, *HLA-DRA*, was at t=5, the midpoint of a typical 10 hour S-phase. Significantly, also, *HSPA1* replicated at a similar time to the early replicating control *PYGM* which was equally enriched in the 1 and 3 hour fractions. The band representing the sample from timepoint t=1, which consisted of ~22% BrdU-positive cells, would be stronger than t=3, which had almost four times as many BrdU-positive cells, had the percentages been similar. As a result, the two loci, *PYGM* and *HSPA1*

would have been enriched mostly in t=1, a result confirming them as very early replicating.

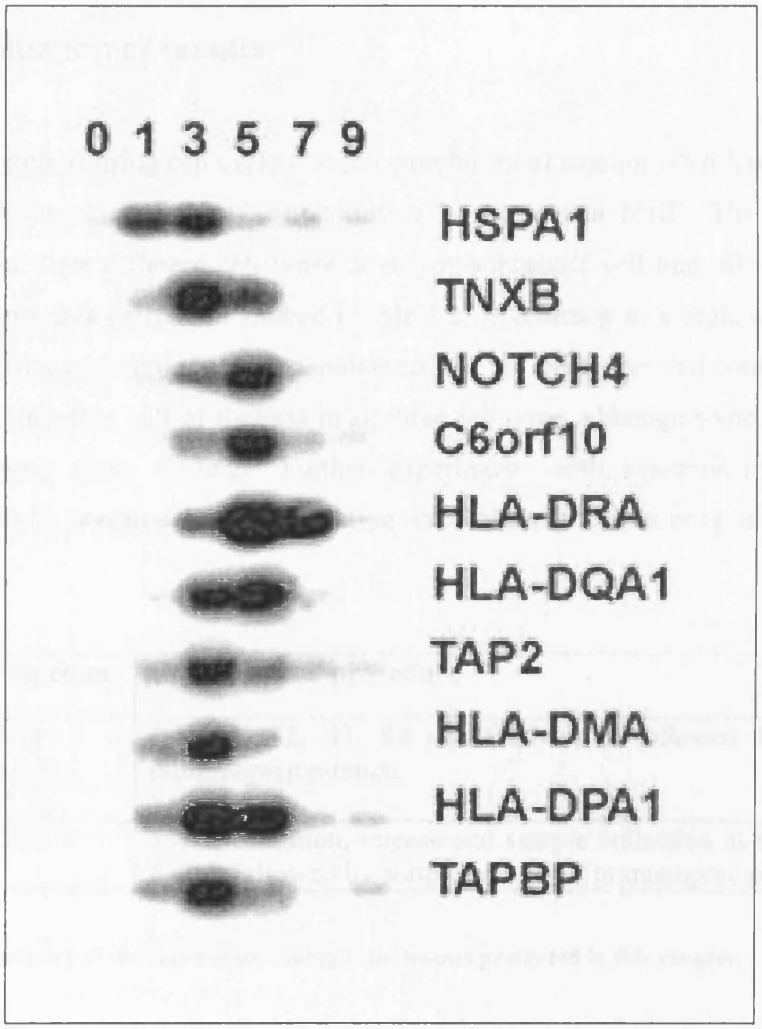


Figure 3.20 Relative abundance of nascent DNA at ten MHC loci for different fixation timepoints post-release from block in HL60 cells.

The MHC loci examined were mostly enriched in the T=1, T=3 and T=5 timepoints indicating that the MHC replicates within the first half of S-phase, which commonly lasts for a total of 10-12 hours. Results shown are typical of at least two independent experiments.

3.3 Discussion

3.3.1 Evaluation of results

Our investigations using cell cycle fraction enrichment of nascent DNA have provided important information for replication timing in the human MHC. The replication behaviour of three different cell types: a B-lymphoblastoid cell line, fibroblasts and myeloid leukaemia cells, was studied (Table 3.2). According to a replication timing assay involving asynchronous cell populations, the ten MHC loci that were examined replicated in the first half of S-phase in all three cell types, although some differences existed among those findings. Further experiments with synchronised myeloid leukaemia cells revealed more information about the replication programme in this region.

Cell type (section)	Experimental procedure
AHB (3.2.1) MRC5 (3.2.2) HL60 (3.2.3)	G1, S1, S2, S3, S4 and G2/M sort followed by BrdU immunoprecipitation
HL60 (3.2.3)	Synchronisation, release and sample collection at t=0 ,1, 3, 5, 7, 9 followed by sorting and BrdU immunoprecipitation

Table 3.2 Summary of the replication timing experiments presented in this chapter.

Firstly, replication timing at ten MHC loci was tested in asynchronous B-lymphoblastoid AHB cells that had been flow-sorted in different cell cycle sub-populations. Enrichment in BrdU-labelled DNA was predominantly found at the S1 sub-population in all ten loci, suggesting that they replicate during an early stage in S-phase. However, a pattern of less enriched bands was seen next to the predominant band for some of the loci.

Three possible explanations can be offered to explain these faint bands. Firstly, as in the case of *HLA-DMA* (Figure 3.9) where the band corresponding to G1 is considerably enriched, it is possible that the locus replicates very early in S-phase and the population of cells where this has occurred have been sorted in the G1 and S1 sub-

populations. Secondly, using the same example, it is also possible that the band corresponding to G1 is present due to the "cross-contamination" of neighbouring populations when these are flow-sorted with respect to DNA content. Thirdly, as in the case of *HLA-DQA1*, the faint bands seen in S2 and S3 might be due to the length of BrdU labelling time and the number of cells collected. Nevertheless, the signal to noise ratio was evidently higher in these compared to previous experiments (data not shown). The band corresponding to the S1 sub-population was significantly more enriched than the bands of other sub-populations.

The pattern of faint bands observed was not always the same from one independent experiment to the next. This observation could suggest that the phenomenon of faint bands is either due to minor differences in the flow-sorting procedure from one experiment to the next in conjunction with the explanations offered above, or other intrinsic imperfections of the methods.

When MRC5 fibroblast cells were used to study replication timing at the ten MHC loci using an identical approach, the phenomenon of faint bands re-emerged. For most loci, enrichment in BrdU-labelled DNA was found in both S1 and S2 in these cells. Notably, there was a discrepancy in the replication timing of the *HBB* locus, which had previously been used as late replicating control in the study of AHB cells. The earlier-than-expected replication at the *HBB* locus was probably due to the higher level of expression in the MRC5 relative to AHB cells, where the locus was found to replicate late in S-phase. This finding is in agreement with reports that the *HBB* gene replicates early when it is expressed and late when it is not (Dhar, Mager et al. 1988).

A myeloid leukaemia cell line, HL60, was the last to be examined. An identical approach to that used to study replication in AHB and MRC5 cells was used initially. The pattern of the bands for HL60 cells was quite similar to the pattern observed with MRC5 cells. For several of the MHC loci tested, enrichment in BrdU-labelled DNA was seen at both bands corresponding to the S1 and S2 sub-populations. HL60 cells were also used to examine replication at these ten MHC loci using a different approach, involving chemical synchronisation with aphidicolin. Aphidicolin stands as a valuable tool for cell synchronisation based on evidence which suggests that it does not interfere with cell viability, S-phase duration, synthesis of dNTPs and of the DNA

polymerases α, β and γ. Methods involving hydroxyurea or thymidine block, on the other hand, have been shown to affect these aspects.

In previous reports, Tribioli *et al.* and Tenzen *et al.* did not directly examine the synchronous progress of the populations in S-phase after release from block, as had been performed by Pedrali-Noy *et al.* (Pedrali-Noy, Spadari et al. 1980; Tribioli, Biamonti et al. 1987; Tenzen, Yamagata et al. 1997). This process is a vital step for establishing the efficiency of synchronisation. Therefore, the protocol used in those studies was modified here with the aim of improving the reliability of the results. Instead of labelling with radioactive nucleotides and BrdU as a means of separating nascent DNA, cells from samples collected at various intervals after release from block were labelled with BrdU and then flow-sorted. This modification facilitated monitoring of the number of BrdU positive cells present in each sample. In these experiments cells were collected at different timepoints to those of Tenzen *et al.*, up to 9 hours after release from block, and included IP and replication timing controls as in the experiments described above.

There was a clear benefit in using aphidicolin synchronisation to analyse replication timing, as seen when comparing Figures 3.17 and 3.20. This is likely a consequence of the fact that the approach involving synchronisation identified the timing of replication with greater precision, due to enrichment in the number of cells with coordinated replication. Admittedly, a perturbation to the physiological cycling conditions of any cell line could introduce artefacts. Nonetheless, the apparent agreement of these data with the findings from other studies which focus on the MHC, reinforce the validity of the results.

Overall, the replication timing for the MHC in all three cell types was found to be in the first half of S-phase. In experiments performed with more homogeneous sub-populations of myeloid leukaemia cells, following chemical synchronisation, it emerged that the ~1.5 Mb under investigation had earlier and later replicating portions. In fact, loci within the classical MHC class II region replicated later than the neighbouring sequences. The earliest replicating locus of those examined here maps within the HSP70 gene cluster, in class III of the MHC, which is a very gene rich region.

3.3.2 Comparisons between studies

To understand better the results of the assays described here, the findings from other studies focusing on the MHC were considered. Do the results from the different studies agree with each other?

The findings presented in this chapter show that the ten MHC loci replicate within the first half of S-phase and suggest an inconsistency with the data from the earlier FISH-based study. Those findings supported the existence of a biphasic replication programme in this region, with some parts of the MHC replicating early and other parts late in S-phase. What are the underlying reasons for the differences between the two studies conducted in the laboratory?

The following explanation can be offered for the apparent disagreement between the two sets of data. The resolution of replication timing between the MHC probes according to FISH is very detailed, with a section of the MHC class II region replicating "late" in S-phase or, more precisely, later than its neighbours. In terms of absolute replication timing, however, the evidence from the biochemical assay indicates that in fact, the loci tested replicate within the first half of S-phase. The FISH data offer an insight into the relative replication programme across the region within each cell type examined. The biochemical data on the other hand offer more precise information into the absolute timing of replication for each cell line but the resolution does not enable detection of details to the levels as seen by FISH.

The FISH assay depends on the discrimination of singlet and doublet signals in populations of unsynchronized cells. How long after replication it takes to detect doublets, and whether or not this time is constant for different sequences, has not been formally addressed. Delayed chromatin resolution of certain replicated loci into doublet signals was previously described when cells were fixed with a method that preserves 3D nuclear protein integrity (Azuara, Brown et al. 2003), but not when fixed with methanol:acetic acid, the method used in our laboratory. Also, poor

hybridization efficiency of the probe can give an artificially low doublet score and thus an erroneous interpretation of late replication (Bickmore and Carothers 1995).

Even in the presence of synchronisation, which essentially relies on the frequency of sample collection intervals, the resolution of the BrdU-IP assay has been limited. Instead of the two-hour intervals used in this study, Tenzen *et al.* who collected samples every hour, had more overlap in the nascent DNA enrichment of consecutive time intervals, since for some of the loci tested two or more bands had similar intensity (Tenzen, Yamagata et al. 1997). Aphidicolin synchronization did not produce a tight, narrow peak of cells passing quickly into and out of each fraction. In fact, it gave rise to relatively broad S-phase sub-populations, which included "contaminating" earlier and later S-phase cells. Flow-cytometric analyses of the fractions collected at different timepoints clearly illustrate this point (Figure 3.18). It was initially expected that the entire cell population would be at the G1/S border after arrest, but in fact a considerable number of cells remained in S-phase and G2/M. Consequently, the results of Tenzen and colleagues are probably affected by a proportion of the HL60 cells taking longer to enter S-phase after two rounds of aphidicolin treatment, likely affecting the findings of that study.

The reported difference of one hour between two primer sets just 16 Kb apart in the GC/AT transition area separating the class II and III regions (Figure 3.1) which had been branded as "precise switching of DNA replication timing" (Tenzen, Yamagata et al. 1997), was likely imprecise. For that report, the GC content had been defined using a commercially available kit (Fukagawa, Sugaya et al. 1995), due to absence of sequence data for the region under investigation. Therefore, the information Tenzen presented was not as detailed and precise in comparison to the sequence information that is available nowadays. Based on current sequence data, the GC content in that region does not change dramatically at the point that Tenzen *et al.* suggested, but has a staggered transition over several tens of Kb with ups and downs. In a previous study by the same group, the following statement was made with reference to chromosomal band boundaries: "Boundaries may be structurally assigned where there are clear GC% transition points, and characteristic signals may be found that punctuate and/or differentiate distinct functions, such as a signal that switches from early to late DNA replication." (Fukagawa, Nakamura et al. 1996). Based on the previous observation of

a small G-subband within 6p21.31 (Senger, Ragoussis et al. 1993), Fukagawa *et al.* proposed that a chromosome band boundary existed between the class II and class III regions. The study by Tenzen *et al.* supports this hypothesis by suggesting the existence of a precise switch in replication timing based on their evidence.

A transition to earlier replication has been identified on either side of the classical MHC class II region. However, even if this segment of the MHC constitutes a separate G-subband, merely the changes in GC content and replication timing with respect to its neighbouring sequences could not define its boundaries to such great detail as has been attempted by Tenzen *et al.* In fact, previous studies examining replication across band boundaries did not identify sharp transitions across those regions (Strehl, LaSalle et al. 1997; Bilyeu and Chinault 1998). Band boundaries would be characterised by a variety of factors to reflect the complexity that defines isochores and chromosomal bands. As discussed in the first chapter of this thesis, features such as sequence composition, GC changes and particular chromatin architectures, contribute to the final picture.

Although the chromosome 6 array study offers useful insights into the replication timing for the entire MHC region (Woodfine, Beare et al. 2005), a couple of issues need to be considered. As discussed by the authors, the flow-sorted G1 fraction contained 2% of cells from early S-phase. Similar observations were also made during the course of this thesis, whereby cells in the G1 phase of the cell cycle, assessed by DNA content, included cells that had incorporated BrdU (refer to Figure 3.4). Inclusion of early S-phase cells in the G1 fraction likely affects the S:G1 ratio of early replicating loci, assigning them a value closer to 1. Another weak point of this study is the variable size of the clones spotted on the arrays. Although the average resolution for the MHC was at 79 Kb, the size of the clones varied from less than 10 to over 100 Kb. Such a considerable difference in size could introduce problems for the analysis of the array data. Between short and long clones the value of the raw signal for the dyes used to label the DNA would differ by as much as a ten-fold.

The findings of the different studies are broadly in agreement. Evidence shows an early to late transition in replication timing between classes III and II of the MHC.

The precise timepoint in S-phase and exact locus in the MHC this transition occurs remain controversial.

3.3.3 What defines replication timing in the MHC?

The timing of DNA replication generally correlates with transcription, gene density and sequence composition. How is the timing affected if a genomic region has a combination of features that individually correlate with either early or late replication? The MHC class II region is an AT-rich isochore that would be expected to replicate late, but it also contains a cluster of coordinately regulated immune-related genes that are highly expressed in antigen-presenting cells and are strongly inducible in other cell types. In these studies the entire MHC was found to replicate within the first half of S-phase. The classical MHC class II region replicates slightly later than the adjacent regions irrespective of gene expression. Why does this region replicate in the early-to-middle period of S-phase, later than the rest of the MHC?

DNA replication in eukaryotes begins at multiple origins across the genome, with each origin firing at a particular time during S-phase (Gilbert 2002). Although the mechanisms that regulate when an origin fires are not well understood, intrinsic genomic features that correlate with replication timing have been identified. These include gene density and DNA sequence composition, since gene dense/GC-rich DNA generally replicates early and gene-poor/AT-rich DNA replicates late during S-phase (Drouin, Holmquist et al. 1994). The clearest manifestation of this relationship is seen on metaphase chromosomes where clustered early- and late-firing origins form replication bands that often coincide with the positions of R- and G-bands (Dutrillaux, Couturier et al. 1976; Holmquist, Gray et al. 1982).

The best-studied correlate of replication timing, however, is transcriptional status. In general, studies on individual genes, chromosomes, and whole genomes show that active genes replicate early in S-phase while silent genes and heterochromatin replicate later (Selig, Okumura et al. 1992; Schubeler, Scalzo et al. 2002; Woodfine, Fiegler et al. 2004; Woodfine, Beare et al. 2005). For example, developmentally regulated genes that play a role in cell fate restriction adopt an early replication

pattern when they become expressed during neural differentiation. In contrast, genes that are associated with maintenance of pluripotency in stem cells shift to a later replication status after commitment (Perry, Sauer et al. 2004). Examples are also seen in imprinted and other monoallelically-expressed genes where the active alleles replicate earlier than the inactive alleles (Ensminger and Chess 2004). This relationship does not hold in all genomic regions, though, such as the highly transcribed and late-replicating region recently identified on chromosome 22 (White, Emanuelsson et al. 2004).

The data presented here show that the entire MHC replicates within the first half of S-phase in all cell populations tested. A large portion of the MHC class II replicates later than the adjacent regions, and there are steep transitions between the earlier and later replicating regions, as was found by the cell synchronisation experiments and a study that examined replication by FISH. However, replication timing does not appear to correlate strictly with gene expression as there was no late replication in untreated MRC5 and in HL60 cells. Rather, the class II region replicates in the early-to-middle period of S-phase.

Replication timing in the MHC is thus likely to be influenced by other features, such as gene density or sequence composition. As a gene-rich region, the MHC would be expected to have early replication. Indeed, the class III region has the highest gene density in the genome, containing an expressed gene every 15 kb or less. The earliest replication was found at this region in agreement with a previous study (Woodfine, Beare et al. 2005). Several lines of evidence suggest that replication timing is also strongly associated with local epigenetic features such as acetylation or methylation of specific histone residues and the presence or absence of certain regulatory protein(s) (reviewed by Lande-Diner and Cedar 2005). Therefore, if chromatin in the MHC is "poised" for rapid transcriptional activation, as suggested by the massive chromatin decondensation that occurs within minutes of exposure to interferon-gamma (Volpi, Chevret et al. 2000), an "open" chromatin architecture might predispose to relatively early replication even in non-expressing cells.

In conclusion, we have shown that replication timing in the entire MHC occurs in the first half of S-phase, and that the class II region replicates slightly later than the

adjacent regions. Our findings suggest that replication timing in the MHC correlates with chromatin in the region being poised for transcriptional activation as an essential component of the cell's defence against pathogens.

Chapter 4

Replication timing in the MHC: a microarray approach

4.1 Introduction

The advent of array technology has been instrumental in the advancement of genome research. With a wide range of applications, from areas like cancer genetics to providing answers for DNA-protein binding affinities, it has also been used to understand replication organisation better.

In 2001, Raghuraman and co-workers investigated the dynamics of DNA replication in the genome of the yeast *Saccharomyces cerevisiae* (Raghuraman, Winzeler et al. 2001). Cells were grown for many generations in medium containing heavy isotopes of carbon and nitrogen, ^{13}C and ^{15}N. They were then arrested before the onset of S-phase and after washing, they were resuspended in isotopically light medium. The DNA synthesised in the ensuing S-phase was labelled with light isotopes. Samples were collected at various timepoints throughout S-phase and following restriction digestion each sample was fractionated by density-gradient centrifugation to separate molecules carrying the different density labels. These were then labelled with biotin and separately hybridised to arrays consisting of short oligonucleotides that overlap open reading frames of the *Saccharomyces cerevisiae* genome. Following data processing, the replication timing for each feature on the array was calculated and the replication profile, which also indicated the position of replication origins and termination zones, was plotted. Surprisingly, no correlation was found between transcriptional activity and replication time for many genes in this organism.

The study on *Saccharomyces cerevisiae* paved the way for genome-wide investigations of replication timing for other eukaryotic genomes. A different approach was used to assess replication timing in a cell line from *Drosophila melanogaster* (Schubeler, Scalzo et al. 2002). An asynchronous population of cells was labelled with BrdU and then cells were flow-sorted according to DNA content

into Early and Late S-phase sub-populations. BrdU-labelled DNA was isolated from each of the two samples. Following amplification and differential dye labelling the two DNA samples were co-hybridised onto microarrays containing 6500 sequences, the majority of which represented expressed sequences from *Drosophila melanogaster*. The key finding was that there was a correlation between transcriptional activity and early replication. However, this did not hold for all sequences tested, since over 20% of the earliest replicating genes were found to be silent and over 20% of the latest replicating genes were expressed. This raises questions over the differences in DNA metabolism between single-celled and multi-cellular organisms.

A second study analysed replication timing at higher resolution for *Drosophila melanogaster* chromosome 2L (MacAlpine, Rodriguez et al. 2004). Cultured cells arrested before the onset of S-phase were labelled with BrdU either during the first or the last hour of DNA replication. BrdU-enriched DNA from each sample was isolated, differentially labelled and the two samples were then co-hybridised onto a custom-made array with nearly contiguous coverage of the chromosome. The data were processed and a smoothing algorithm was applied to generate a replication timing profile that was similar to that of Schubeler et al (Schubeler, Scalzo et al. 2002). Most significantly, the study revealed a strong correlation between the local density of RNA Pol II in broad regions (>100 kb) along the chromosome and replication timing.

Several investigations of replication timing have been performed on human cells. Analysis of human chromosome 22 using a tiling path array with >90% coverage, was performed by White and colleagues. They employed a similar approach to Schubeler *et al.* to examine replication in different cells types, a fibroblast and a lymphoblast cell line. Although many genes are differentially expressed in these two cell types, the replication timing profiles were virtually identical. As mentioned in Chapter 3, some late replicating genes on this chromosome were found to be highly expressed.

Taking a different approach, Jeon released HeLa cells from G1-S block, pulse-labelled them with BrdU at 5 two-hour intervals, representing the entire duration of S-phase, and then performed restriction digestion (Jeon, Bekiranov et al. 2005). BrdU-

enriched DNA was then isolated from each two-hour sample by two rounds of density-gradient centrifugation to separate molecules of different densities, followed by immuno-separation. Each DNA sample was co-hybridised with genomic DNA, after differential labelling, onto a cDNA array covering chromosomes 21 and 22. The replication profile in Jeon's study was compiled from each two-hour sampling component. While early replication was found to correlate with high gene density and high transcriptional activity, about 60% of the sequences assayed replicate through the four quarters of S-phase without enrichment in any particular time period. It remains to be established whether this is related to the aneuploid and malignant characteristics of the cells.

The relatively simple approach of Woodfine *et al.* has been described above. They co-hybridised differentially labelled S-phase and G1 DNA to a genome-wide array at 1 Mb resolution, and to tiling path arrays of chromosomes 6 and 22q at resolutions of approximately 94 Kb and 78 Kb, respectively (Woodfine, Fiegler et al. 2004; Woodfine, Beare et al. 2005). The key finding from these studies is a general correlation between replication timing, GC content, gene density and repeat sequence density.

The various studies described here were able to generate quantitative data of replication timing across large genomic regions or even entire genomes. This is an important advantage for analysing relationships with genomic components such as gene richness or epigenetic attributes like histone modifications. The use of microarray technology in the study of replication is not devoid of weaknesses, as discussed later. Nonetheless, it offers useful insights on the mechanism and factors influencing replication timing.

In the latter part of this project, a high resolution tiling path array covering the entire human MHC region became available, providing a powerful tool for the analysis of replication timing. The array, which was constructed at the Wellcome Trust Sanger Institute in the laboratory of Dr Stephan Beck, had very few coverage gaps and included both coding and non-coding sequences at a resolution of ~2.5 Kb. This chapter describes the use of the array, taking a similar approach to Woodfine *et al.* (Woodfine, Beare et al. 2005), for analysing replication timing. The lymphoblastoid

cell line AHB which was studied previously at 10 different MHC loci using southern hybridisations (Chapter 3), was used for this study.

4.2 Results

An asynchronous population of AHB lymphoblastoid cells was flow-sorted based on DNA content. Two cell cycle sub-populations were collected, G1 and S-phase (Figure 4.1). Equal amounts of DNA from each sample were differentially labelled with fluorescent dyes. Before hybridisation of G1 (control) and S-phase (test) DNA samples to the array, a pre-hybridisation step with Cot1 DNA was performed in order to block repetitive sequences (Woodfine, Fiegler et al. 2004; Woodfine, Beare et al. 2005). To enable reliable comparisons between the findings of the two studies, the principles described therein were adhered to.

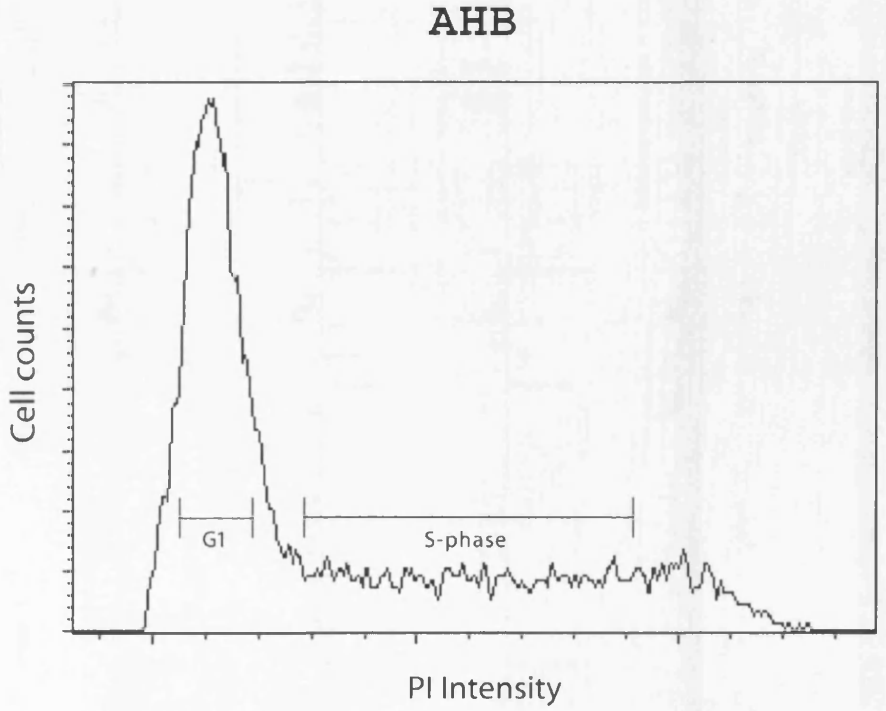


Figure 4.1 Typical FACS profile of AHB cells used in microarray experiments.

Two different sub-populations, G1 and S-phase, were flow-sorted according to DNA content from an asynchronous population of AHB lymphoblastoid cells. The two cell cycle fractions are demarcated by a pair of vertical lines representing the flow-sorting gates used for these experiments. Gaps were left between the two sets of gates to reduce overlap between the two fractions.

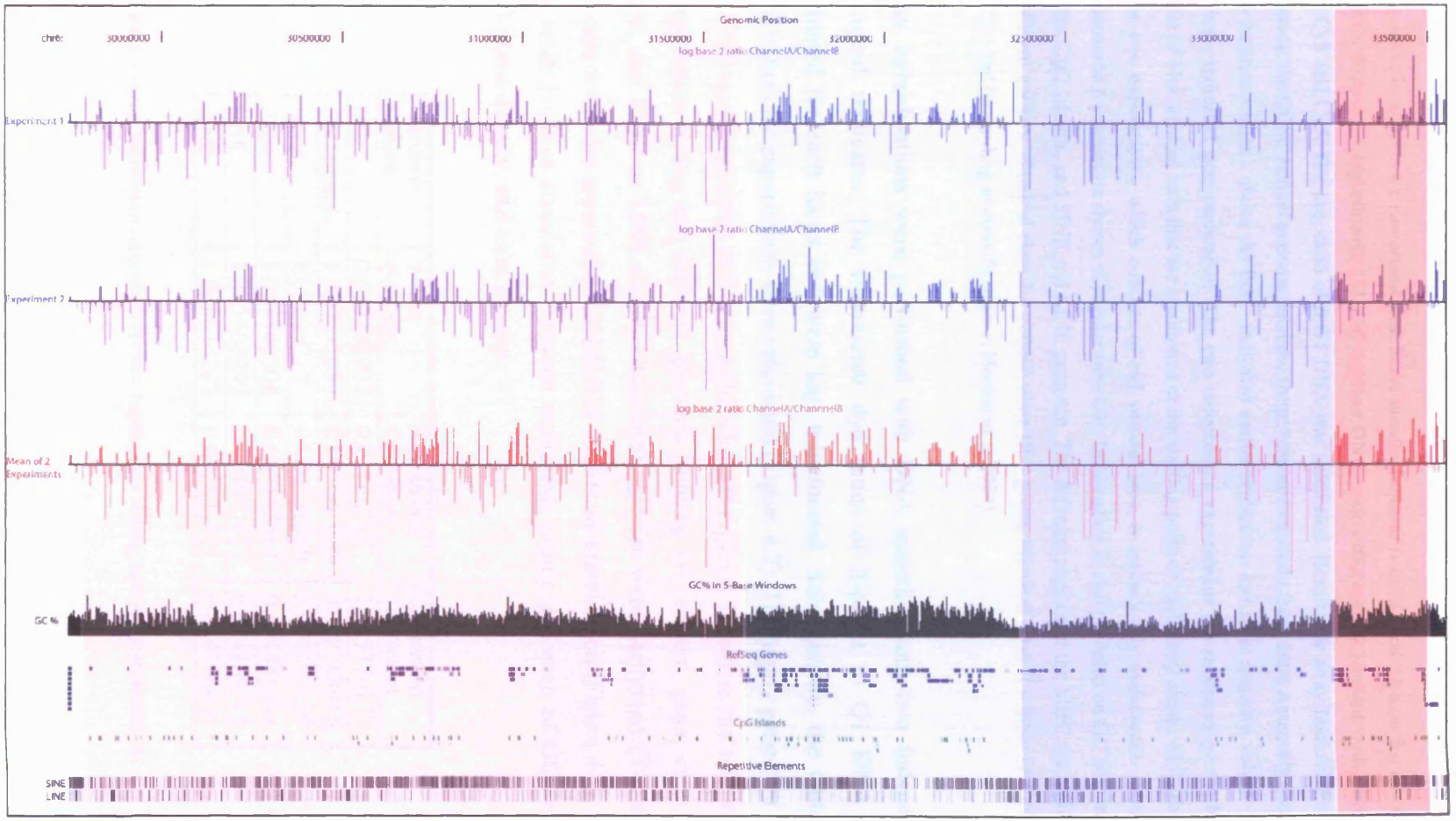

Genomic Position
chr6:
30000000
30500000
31000000
31500000
32000000
32500000
33000000
33500000
log base 2 ratio ChannelA/ChannelB
Experiment 1
log base 2 ratio ChannelA/ChannelB
Experiment 2
log base 2 ratio ChannelA/ChannelB
Mean of 2 Experiments
GC% in 5-Base Windows
GC %
RefSeq Genes
CpG Islands
Repetitive Elements
SINE
LINE

Figure 4.2 (above) Assessing replication timing in the MHC with microarrays.

Array hybridisations were performed with DNA samples isolated from G1 and S-phase fractions sorted in two independent experiments. G1 and S-phase DNA were differentially labelled with fluorescent dyes Cy3 and Cy5. The $\log_2$ ratio of S:G1 DNA was calculated. Results for array features were then plotted on the UCSC human genome database (http://genome.ucsc.edu, May 2004 Assembly), such that more positive values - above the line - indicated earlier replication and more negative values - below the line - indicated later replication. The two independent experiments are annotated on the left and consist of blue vertical bars that are positioned at the starting point of each array clone, while the mean of the two experiments, which consists of red vertical bars, is immediately underneath. At the top, chromosome 6 position is shown at regular intervals. Information is also provided on GC content, gene density, CpG islands, and SINE and LINE presence. The different regions of the MHC are highlighted in different colours: classical class I – yellow; class III – green; classical class II – blue; and extended class II – red (according to classification in Horton et al 2004).

Array hybridisations were performed with DNA material from two independent biological replicates. The fluorescent dye ratios of S-phase to G1 DNA were calculated for each locus and were $\log_2$ transformed. Subsequently, the combined mean values for experiments were calculated (Figure 4.2). The highly gene-rich MHC class III region replicates earlier than the adjacent regions. To assess the relationship of replication timing and different genomic features: GC content, gene, exon, CpG island, and SINE or LINE density; statistical analyses were performed (Table 4.1). The data were also represented graphically for visual examination (Figure 4.3). There is a weak positive correlation between replication timing and each of GC content, CpG island density, and exon density.

Genomic feature	Y-Axis Intercept	Regression Coefficient	Correlation Coefficient
GC	0.553703037	0.01146745	0.115376173
CpG islands	1.090631687	0.007041608	0.118743016
Gene	1.07105712	0.000939509	0.074577199
Exon	1.097545508	0.008543861	0.124481344
LINE	1.085463293	0.001975027	0.074469715
SINE	1.063634293	0.002831921	0.084716409

Table 4.1 Linear regression analysis between replication timing and genome statistics.

The relationship between mean ratio values for the clones present on the array and their content level for six different genomic features was examined. GC content, CpG island density and exon content are three features with weak positive correlation to replication timing.

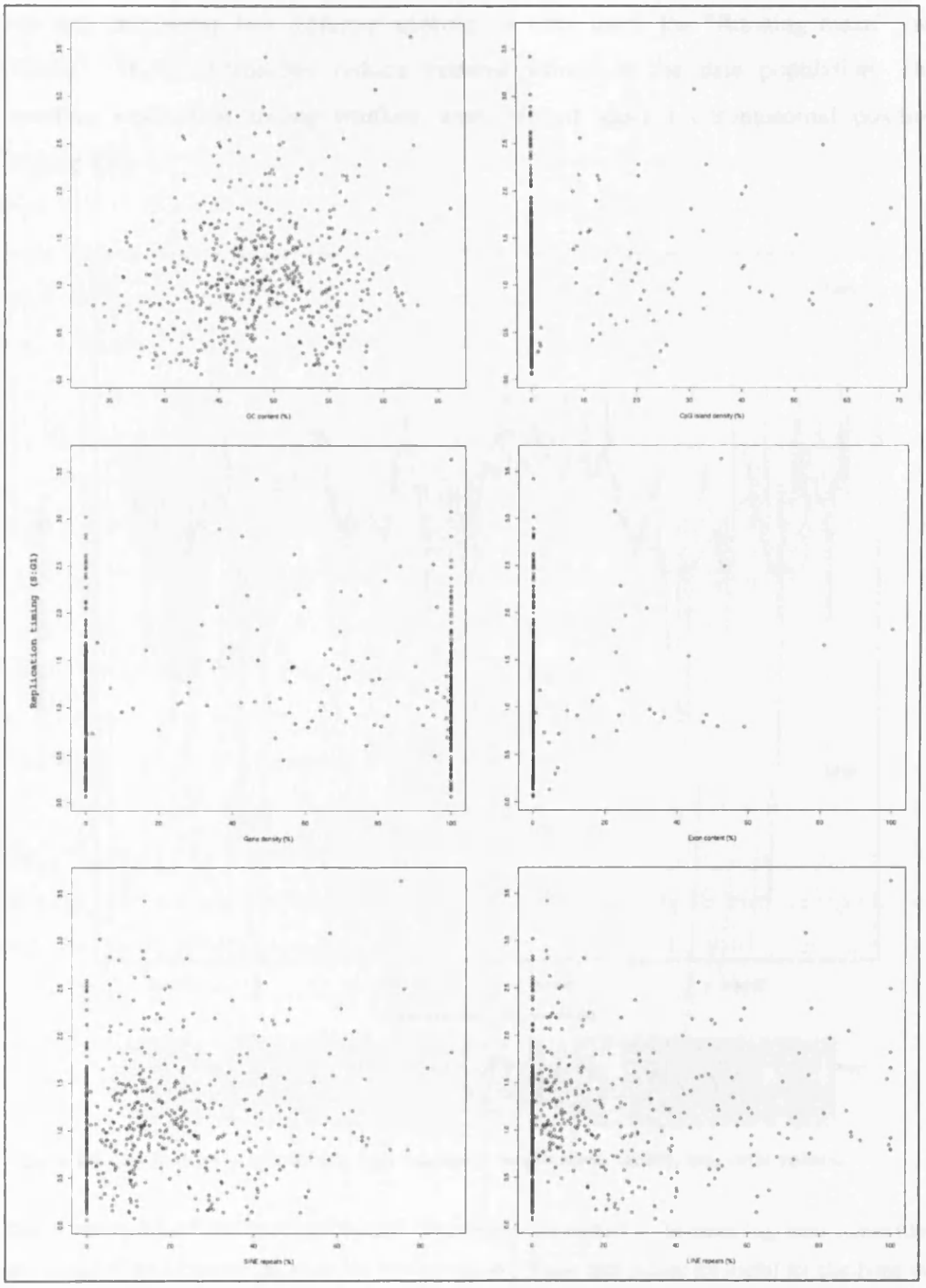


Figure 4.3 Graphical representation of the relationship between replication timing and genome features.

Mean ratio values (S:G1) of two experiments for clones represented on the array were plotted against the content or density of each of six genome features. (Please note that the ratios in these figures are not $\log_2$ transformed.)

For data smoothing two different approaches were used: the "Running mean" and "Loess". These approaches reduce extreme values in the data population. The resulting replication timing profiles were plotted against chromosomal position (Figure 4.4).

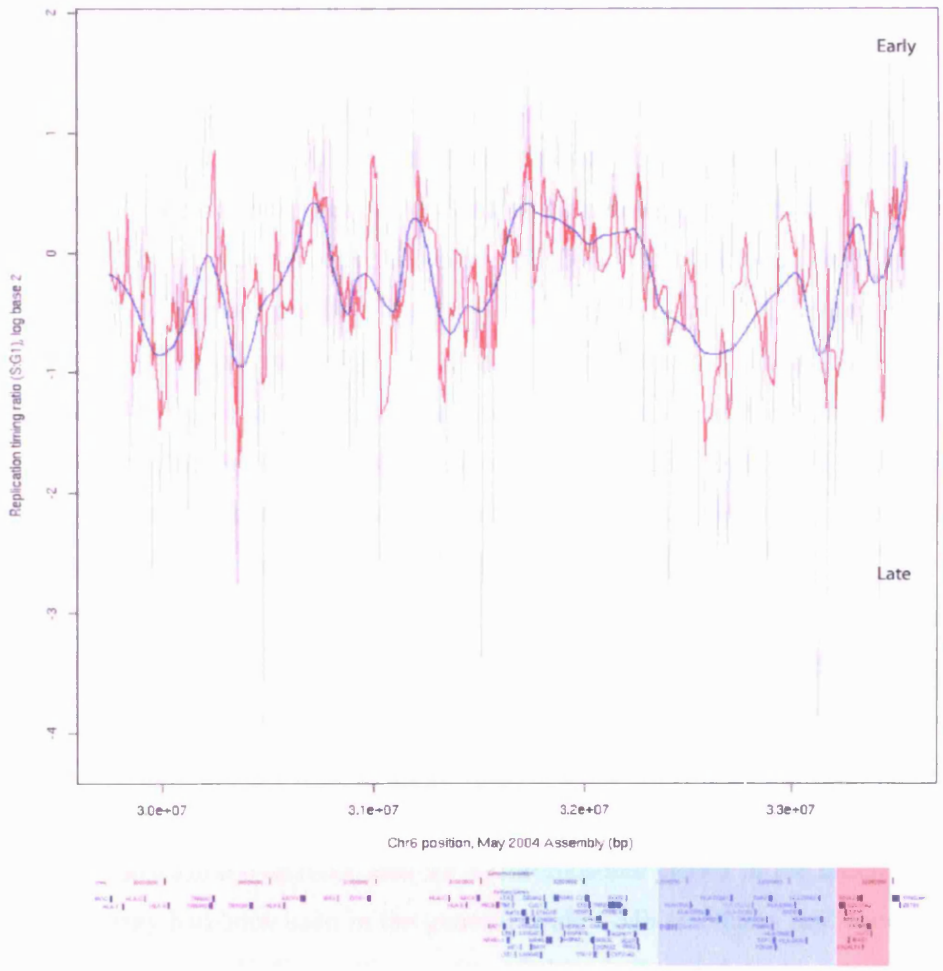


Figure 4.4 Application of smoothing approaches to Replication timing, log_2 ratio values.

The "Running mean" (red line) and "Loess" (blue line) were applied to the mean log_2 ratio values (thin grey lines) of the clones on the array for data smoothing. These approaches are useful for clarifying the trends in the replication timing profile as they reduce the levels of extreme values. Generally, regions with positive log_2 ratio values replicate earlier whereas those with negative values are later replicating. The "Running mean" approach has retained more fluctuation in the plot than the "Loess" curve. The latter clearly defines the various regions as earlier or later replicating. Just to the right of the middle, the class III region features a plateau-like replication profile, with earlier replication than the adjacent sequences. Two early replicating regions are present in class I and one in the extended class II region of the MHC. Background highlighting as per page 49.

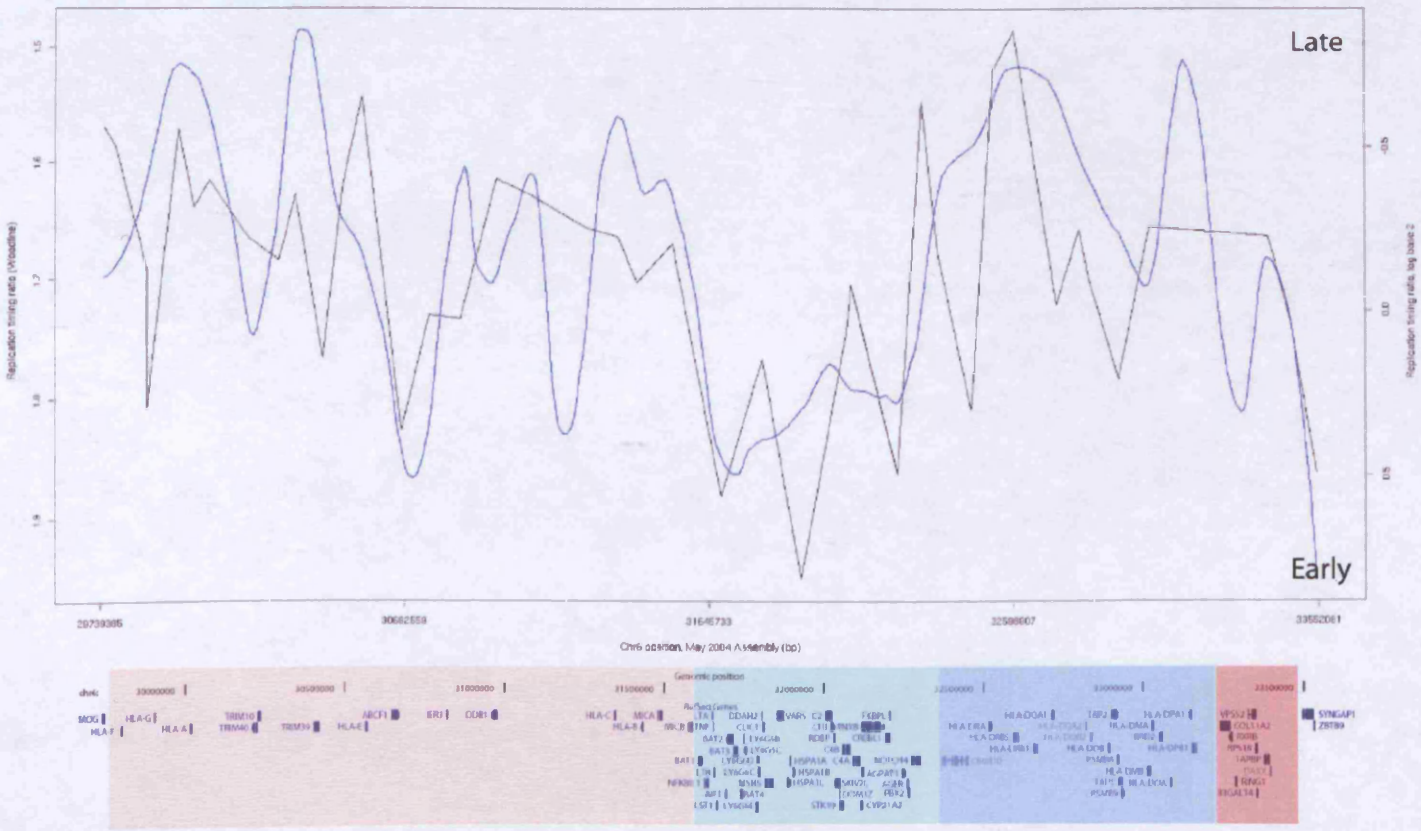
4.3 Discussion

The Woodfine study using the tiling path microarray of chromosome 6 with an overall resolution of 94 Kb concluded that the MHC (resolution of 78 Kb) contains the earliest replicating DNA on the entire chromosome (Woodfine, Beare et al. 2005). To compare the findings of this thesis with those of Woodfine, the smoothened array data presented above were plotted on the same horizontal axis as the data from Woodfine *et al.* (refer to Figure 3.2). Due to the different positions and span of the clones it was not possible to provide statistical evidence for potential relationships between the two sets of data. The smooth curve produced by the "Loess" approach is broadly similar to Woodfine's replication profile (Figure 4.5-A). In general, the "Running mean" approach, which gives a rugged profile, is also in agreement with the published profile (Figure 4.5-B). However, as it contains more peaks and troughs, it discloses more information, extending the published data on the progress of replication in the region. Overall, the data presented here are thus in agreement with the findings reported by Woodfine (Woodfine, Beare et al. 2005).

The S:G1 ratios for the Woodfine study (Woodfine, Beare et al. 2005) that covered the entire chromosome 6 were normalised to the average replication timing for the chromosome, that had previously been reported from the replication timing analysis of the whole genome (Woodfine, Fiegler et al. 2004). This was facilitated by the fact that 155 chromosome-6-specific and 38 X-chromosome clones in the chromosome 6 tiling path array had been used in the genome-wide 1 Mb resolution replication study. In contrast to those findings, the results of this study need to be examined and interpreted in relative terms. These results could only assume an absolute value if normalised with a genome-wide study.

The inferences drawn from figure 4.5 advocate the immense importance of array resolution. The increase in the level of detail achieved for the replication profile is easily justified by the ~40-fold resolution upgrade from the lower resolution array used by Woodfine (Woodfine, Beare et al. 2005) to the array that was used here. In addition, these findings emphasise the significance of data processing and smoothing.

A)


Late
Early
Replication timing ratio (Woodfine)
Replication timing ratio, log base 2
29739385
30682559
31646733
32588907
33552081
Chr6 position, May 2004 Assembly (hg)
Genomic position
chr6:
30000000
30500000
31000000
31500000
32000000
32500000
33000000
33500000
Refseq Genes
MOG
HLA-G
HLA-F
HLA-A
TRIM10
TRIM40
TRIM39
HLA-E
ABCF1
ER1
ODB1
HLA-C
MICA
HLA-B
TNF
CLIC1
DAT2
RDBP
BAT3
HSPA1A
C4A
HSPA1B
ACAP1
BAT1
MSH5
HSPA1L
BAT4
LST1
STK19
CYP21A2
HLA-DOA1
TAP2
HLA-DPA1
HLA-ERA
HLA-DRB5
HLA-DRA
HLA-DPB1
PSMB8
TAP1
HLA-DMA
PSMB9
VPS52
GOLGA2
RGS5
RPL14
TAPBP
TAGY
RING1
BTGAL34
SYNGAP1
ZBTB9

B)

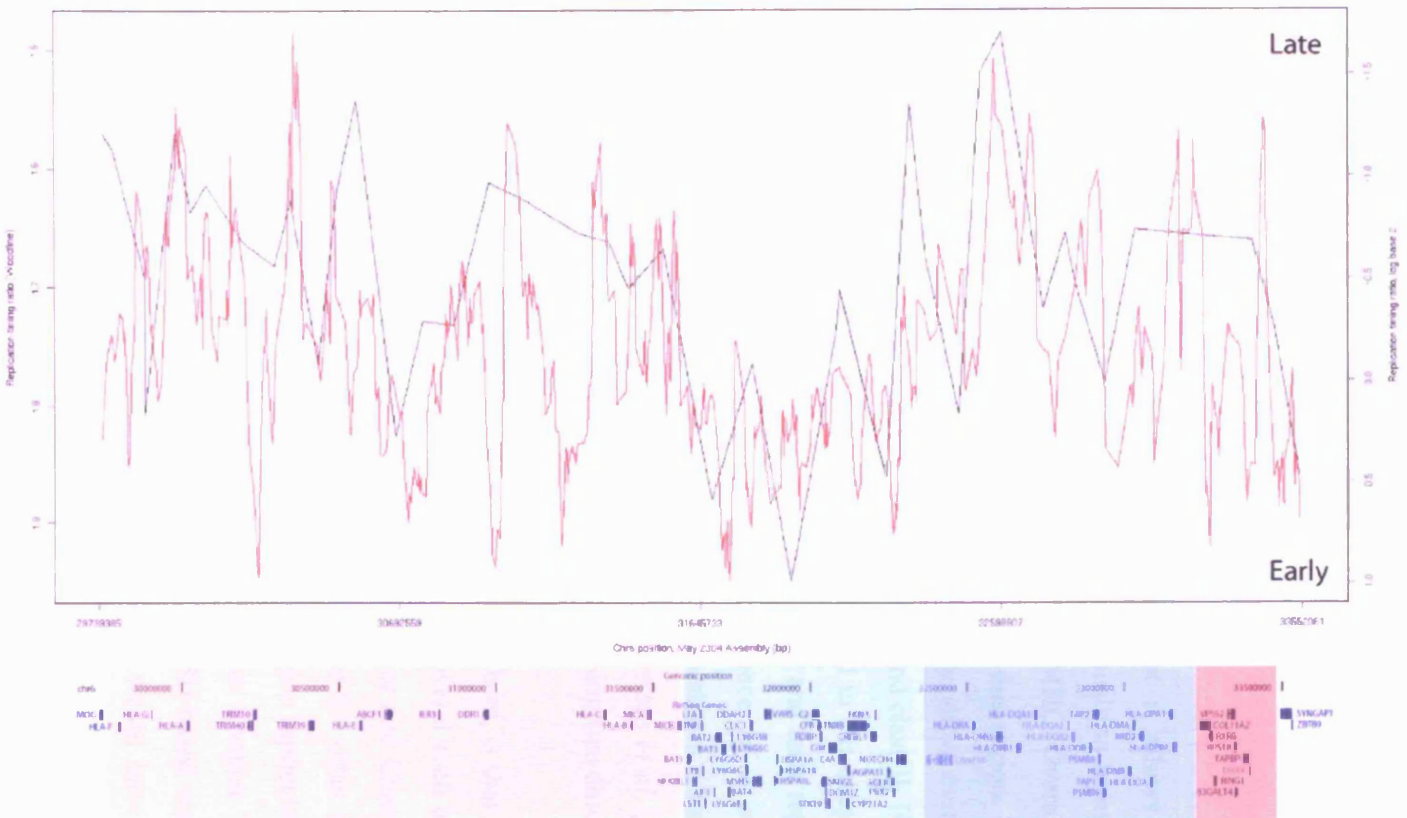


Figure 4.5 (above) Replication timing in the MHC: a comparison between two studies

The replication profile with data smoothing analysis: A) The "Loess" curve (blue line) or B) the "Running mean" (red line); is shown together with the findings of Woodfine *et al.* (black line). For the purposes of this comparison, in order to comply with the convention used in Chapter 3, the vertical axis of S:G1 ratio for the experiments described in this chapter was inverted (Appendix F). Please note that the vertical axes are different. The scale on the two vertical axes was optimised to fit the figure. Background highlighting as per page 49.

The primary advantage in this high resolution study is that it allows reliable comparisons to be made between the different sub-regions. A transition from early to later replication is apparent between class III and class II of the MHC. Overall, this is in agreement with the studies that have examined replication in this region. Both the BrdU-IP assay at ten different MHC loci and the FISH study show evidence of transition from early to later replication between MHC class III and class II. However, their resolution is not as high as that of the array assay presented in this chapter. It is not entirely clear how sharp is the transition between the earlier and later replicating positions on that segment. Figure 4.5-B suggests the presence of a step-wise transition. Intriguingly, a particular segment at around 32.4 Mb shows considerable replication delay. This overlaps with the region examined by Tenzen *et al.* (Tenzen, Yamagata et al. 1997), where a "precise switch" in replication timing was described.

A limitation for array-based assays, including the one described here, is that they are not able to distinguish regions with asynchronous replication. Several such loci have been identified in various parts of the human genome (Ensminger and Chess 2004). For diploid normal human cells, if one allele replicates early and the other one late, then the S-phase pool of DNA will indicate that the particular region replicates in mid S-phase. For the same scenario, the BrdU-IP assay described in Chapter 3 would indicate peaks in two S-phase sub-populations. Similarly, a FISH-based replication assay (refer to section 3.1 for explanation of the technique) would have a high proportion of nuclei, each with a singlet and a doublet.

As previously discussed in Chapter 3, overlap between the flow-sorted cell cycle sub-populations is another potential weakness of the method used here. Despite the fact

that generous gaps were allowed between the G1 and S-phase fractions, those could not be very wide as it would result in the removal of very early replicating sequences.

In light of the findings in this chapter it is appropriate to reconsider the factors affecting replication timing in the MHC. The statistical analysis performed on the mean array values shows a weak positive correlation between GC content and replication timing, such that higher GC content would be associated with earlier replication. A positive correlation also exists between exon content, as well as CpG island content and replication timing. In general, exon-rich regions where one or more genes are present have relatively high GC-content. Such regions are known to have a high occurrence of CpG di-nucleotides, mainly in the form of CpG islands. Taken together, these results suggest that within the human MHC, regions with higher gene density have earlier replication. This conclusion is in agreement with previous studies that made similar discoveries. Nevertheless, in order to establish a definitive link between replication timing and gene expression in the MHC, further experiments are required.

Results in this chapter extend previous findings from this thesis and elsewhere that the classical class II region of the MHC replicates later than adjacent sequences, even in cell-types which actively express the residing genes. As mentioned earlier, highly-transcribed late replicating regions were found by the study of White *et al.* on chromosome 22 (White, Emanuelsson et al. 2004). The phenomenon of an actively expressed genomic segment with later replication is therefore not unique to the MHC region. According to Jeon *et al.* (Jeon, Bekiranov et al. 2005), early replication corresponds to higher gene expression. It remains to be established exactly how late the "later-replicating" regions in the MHC actually replicate as well as the molecular mechanisms involved in differential timing of origin firing.

Chapter 5

Initiation of Replication in the MHC

5.1 Introduction

The replication timing of a region is defined by its proximity to an origin of replication and the time within S-phase that this is activated. Only a handful of replication initiation sites have been mapped and characterised so far in the human genome. The best studied of these are located close to genes whose genomic context and product function have been at the epicentre of scientific attention for years. Mammalian replication origins are frequently located near to promoter regions of genes (Vassilev and Johnson 1990; Giacca, Zentilin et al. 1994; Taira, Iguchi-Ariga et al. 1994; Keller, Ladenburger et al. 2002; Ladenburger, Keller et al. 2002; Paixao, Colaluca et al. 2004). Does this mean that gene poor areas of the genome have fewer origins than gene rich regions? In the absence of a definitive sequence signature for origin recognition, the human genome has been awaiting for the emergence of an approach that will drive research in this field forward.

Several labour-intensive techniques have been employed in an attempt to identify sites of replication initiation. For instance, one method is based on the isolation of leading strand nascent DNA, which is then probed with radioactive oligonucleotides. These synthetic molecules correspond to sequences on either of the two strands and enable identification of replication direction. The origin of bidirectional replication can therefore be predicted (Aladjem and Wahl 1997; Verbovaia and Razin 1997). Another method, which is predominantly used for smaller genomes, involves two-dimensional gel electrophoresis of genomic DNA from synchronised S-phase cells. Radioactively-labelled probes that cover a region of interest are then used to identify replication intermediates. Those probes closest to the origin are expected to exhibit a particular pattern when exposed to film (Huberman 1997).

The FISH analysis that was performed in the Human Cytogenetics Laboratory Cancer Research UK to assess replication timing, gave good temporal resolution revealing a consistent pattern of peaks and dips in all three cell populations examined (Figure 3.3). Previously in this thesis, *HSPA1L* was identified as the earliest replicating locus of the ten MHC loci studied. Intriguingly, a FISH probe encompassing this gene was also found to be the earliest replicating in the three cell types tested. An earlier investigation reported the presence of a replication origin within the heat-shock protein cluster (Taira, Iguchi-Ariga et al. 1994). Therefore, these data suggest that the dips in the FISH assay graph may coincide with the positions of replication origins. It would be possible to examine the entire region covered by the FISH probe for sites of replication initiation using PCR or, as described above, leading strand analysis. However, examination of a genomic region covered by a cosmid at short intervals would require the design and use of several oligonucleotides with short gaps between them.

As described in Chapter 3, Tribioli *et al.* used a chemical synchronisation approach to arrest cells at the G1/S border, and isolate DNA which replicates at the onset of S-phase (Tribioli, Biamonti et al. 1987). This led to the identification of a highly transcribed novel lamin gene, *LMNB2*, on chromosome 19 (Biamonti, Giacca et al. 1992). Further analysis of the genomic locus in association with a pioneering technique allowed the group to map an origin of replication in great detail (Giacca, Zentilin et al. 1994; Abdurashidova, Deganuto et al. 2000). Intriguingly, this origin was found to be active in cells of different derivation suggesting a similarity in the replication programme of those cell types (Kumar, Giacca et al. 1996).

The technique developed by Giacca and his colleagues was based on the isolation of ssDNA within a specific size range from asynchronous cell populations. The isolated material corresponds to nascent DNA from leading strand synthesis. The smaller Okazaki fragments on the lagging strand are avoided due to their shorter length. The pool of nascent DNA molecules was then used as template in competitive PCR reactions with primer pairs covering the region of interest. The most abundant template for that region could then be identified as the origin of replication (Giacca, Pelizon et al. 1997). The rationale of the assay is as follows: The cells entering S-phase activate the lamin B2 (LB2) origin and replication forks progress bi-

directionally away from that site. In an asynchronous cell population other cells enter S-phase at a later time and some others even later on. Those cells which have advanced into S-phase the most would contain forks that have progressed further away compared to those cells that entered S-phase later on and thus have fired the LB2 origin more recently. Consequently, there will always be more nascent DNA and therefore PCR template close to the initiation site than further away.

Other groups have identified replication origins using a similar method for nascent DNA isolation, but different approaches for subsequent analysis. Examples include the origins near the gene encoding the aldehyde dehydrogenase 4A1 (*ALDH4A1*) precursor on chromosome 1 (Brylawski, Cohen et al. 2004) and the hypoxanthine-guanine phosphoribosyl-transferase (*HPRT1*) gene on the X chromosome (Cohen, Brylawski et al. 2002).

The characterisation of replication initiation sites is essential for a comprehensive understanding of replication organisation in the human MHC. The availability of a high resolution MHC tiling path array, constructed by Dr. Stephan Beck at the Wellcome Trust Sanger Institute, facilitated the progression of this project as it offered a fast and reliable means for examining replication initiation sites.

5.2 Results

To identify replication initiation sites within the MHC, single-stranded nascent DNA was first hybridised to the tiling path array covering the MHC. Three asynchronously growing B-lymphoblastoid cell populations were used: PGF– one population, and AHB – two populations. Using the separation pattern of the DNA standard as sedimentation velocity reference, fractions 16-20 containing ssDNA within a size range approximately equivalent to 900 and 1700 bp were pooled together (Figure 5.1), as this had previously been shown to be enriched with nascent DNA (Cohen, Brylawski et al. 2002; Brylawski, Cohen et al. 2004).

To confirm that the fractionated DNA was enriched in replication initiation sites, nascent DNA abundance was examined at two genomic sites, near the *LMNB2* and

MYC genes that had previously been identified as origins of replication. The origin-containing regions were called LB2 and Myc11. Abundance of nascent DNA in these two positive control regions was compared with proximal loci for which no significant enrichment had been reported, called LB2C1 and Myc1 (Giacca, Zentilin et al. 1994; Abdurashidova, Deganuto et al. 2000; Tao, Dong et al. 2000), The genomic distances between the origin-containing loci and the respective negative controls were 4 and 6 Kb, respectively.

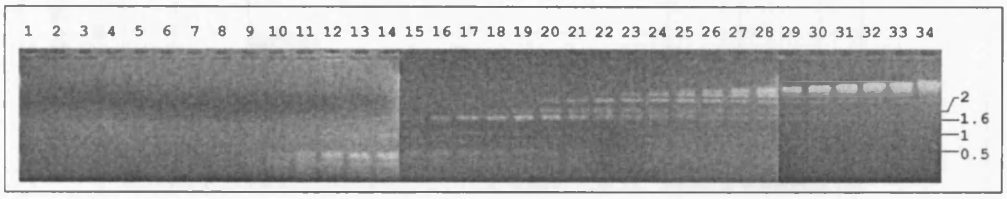
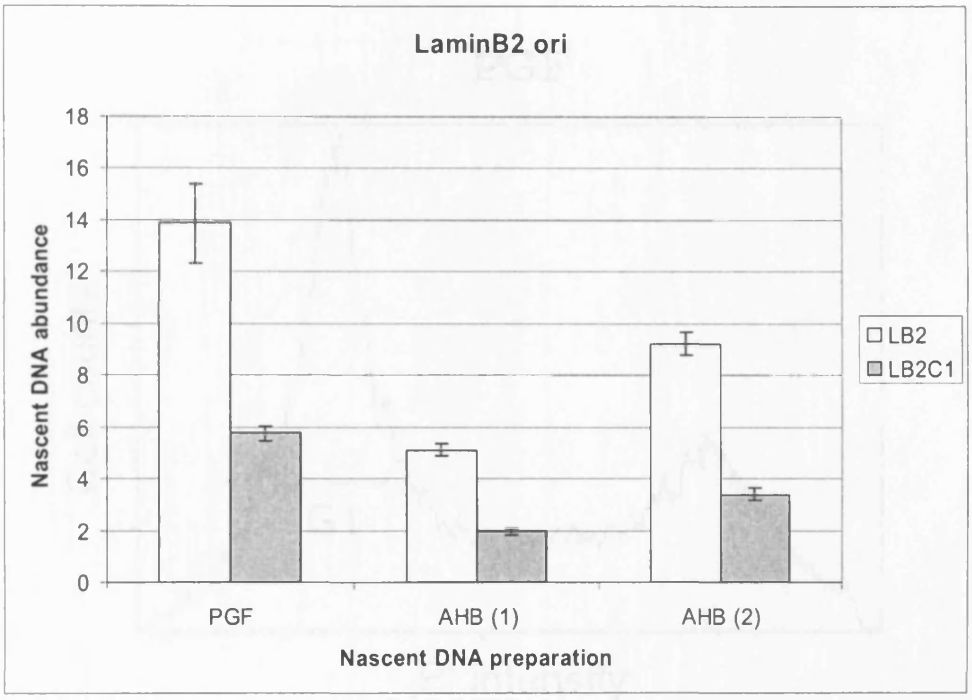


Figure 5.1 Fractionation of 1 Kb DNA marker on sucrose gradient.

Three preparations of ssDNA were run on 5-20% sucrose gradients. A 1 Kb marker was run parallel to those samples under the same conditions. Numbers 1-34 indicate samples corresponding to each 1 ml fraction collected. The numbers on the right (0.5, 1, 1.6 and 2) indicate the size in Kb. Fractions 16-20 from each preparation were pooled together and used in subsequent experiments.

SYBR Green quantitative real-time PCR was then performed and a standard curve produced for each primer pair from reactions with known amounts of genomic DNA. A considerable difference in nascent DNA enrichment was observed between the origin and control regions (Figure 5.2), indicating that these samples were suitable for further analysis.

To identify sites of replication initiation within the MHC region, similar quantities of nascent (test) and G1 (control) DNA were co-hybridised onto arrays, after labelling with fluorescent dyes. The control DNA was obtained from flow-sorted cells at the G1 phase of the cell cycle (Figure 5.3 for PGF; Figure 3.6 in Chapter 3 for AHB). To block repetitive sequences pre-hybridisation with human Cot1 DNA was performed prior to application of the Cy3 and Cy5-labelled DNA samples. Highly abundant SINE and LINE elements can potentially exhibit unspecific binding and thus provide misleading information if this step is omitted.



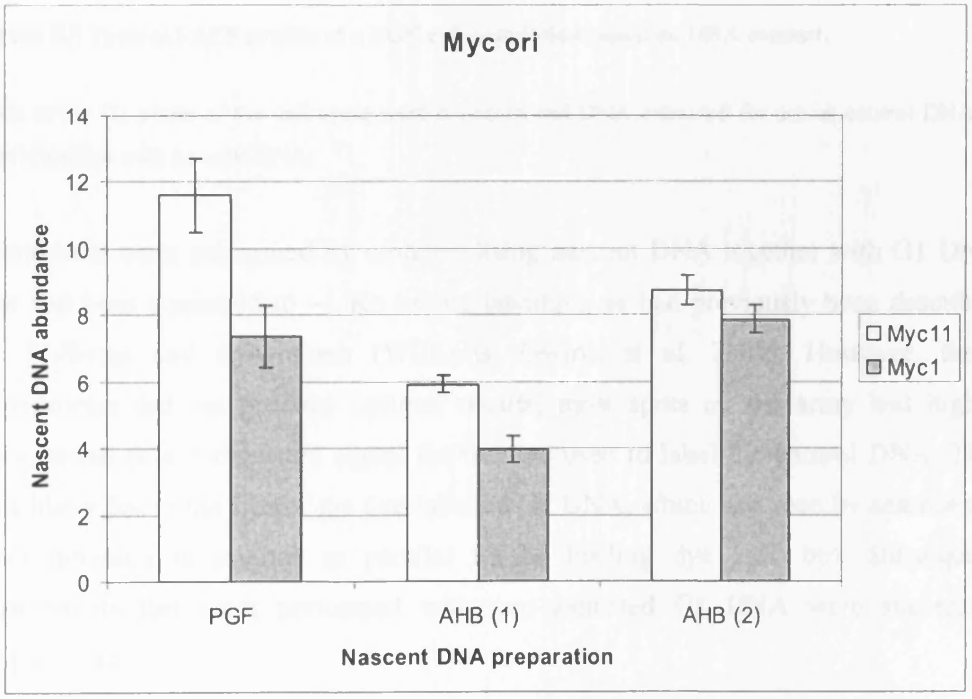


Figure 5.2 Nascent DNA enrichment at the *LMNB2* and *MYC* loci.

Real time PCR experiments were carried out with primers spanning known origins of replication and their respective negative control sites. Bars represent the mean of three experiments with standard deviation. The greatest difference in enrichment for the three samples was observed between the Lamin B2 origin and its control site.

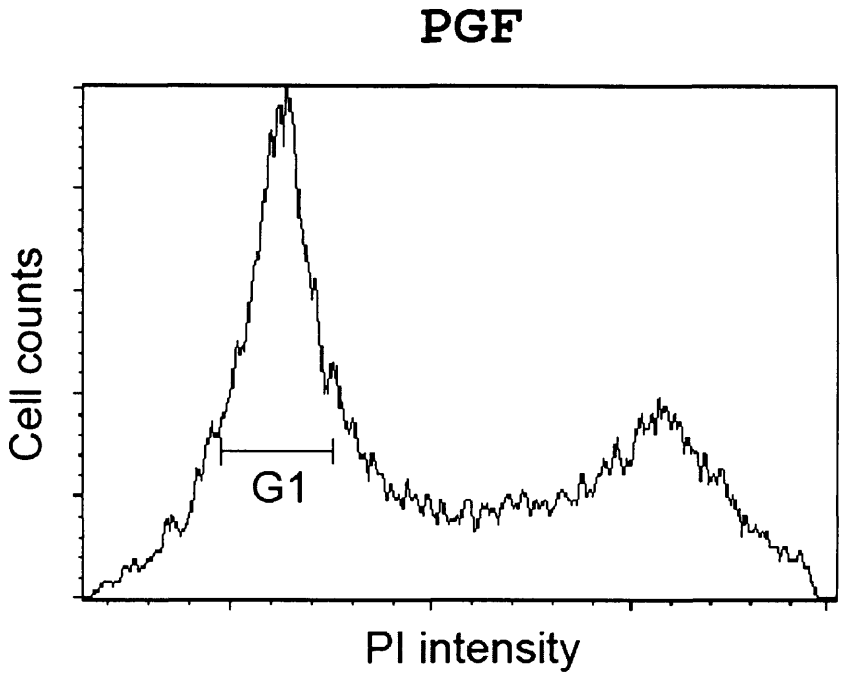


Figure 5.3 Typical FACS profile of a PGF cell population based on DNA content.

Cells in the G1 phase of the cell cycle were collected and DNA extracted for use as control DNA in hybridisations with nascent DNA.

Initial tests were performed by co-hybridising nascent DNA together with G1 DNA that had been sonicated to ~1 Kb before labelling, as had previously been described by Williams and co-workers (Williams, Gwirtz et al. 2004). However, those experiments did not produce optimal results; most spots on the array had higher background than foreground signal for the dye used to label the control DNA. This was likely due to the size of the dye-labelled G1 DNA, which was seen by agarose gel electrophoresis to migrate in parallel to the loading dye (~50 bp). Subsequent experiments that were performed with non-sonicated G1 DNA were successful (Figure 5.4).

An unexpectedly large number of clones showed enrichment of nascent DNA in the three samples tested (Figure 5.4). Several regions were enriched in nascent DNA across several clones and almost up to 100 Kb. One such example is the region between 30.95 - 31.05 Mb, where enrichment is observed in as many as 30 clones.

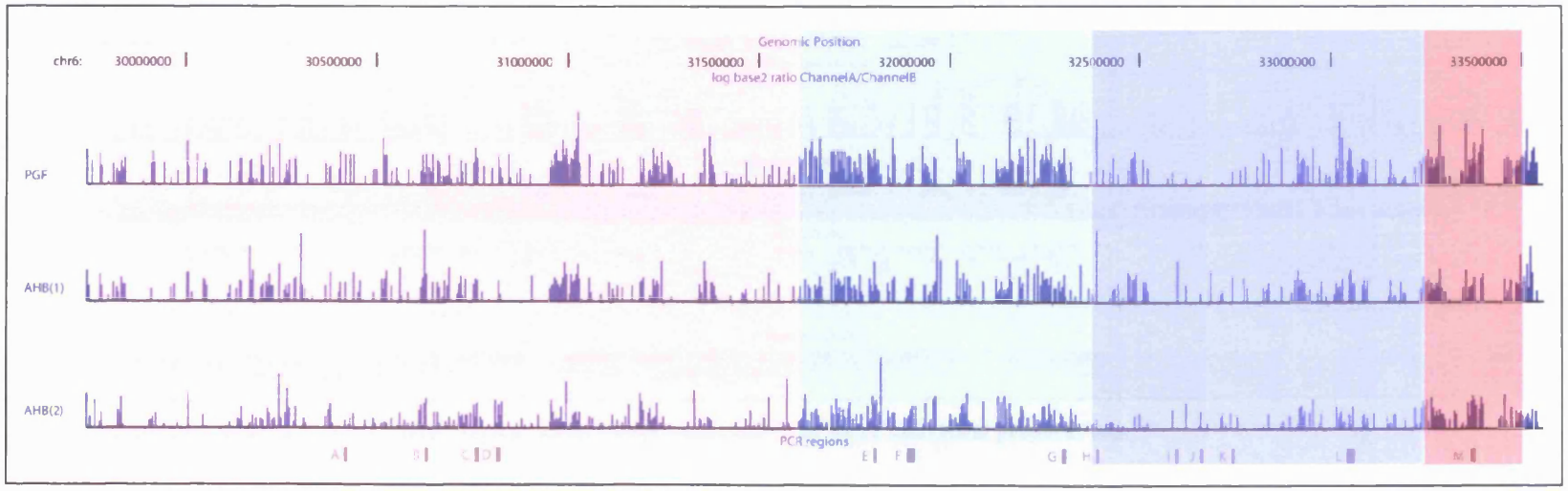


Figure 5.4 Nascent DNA abundance in the MHC region.

Three different preparations of nascent DNA that were isolated from PGF and AHB cells were used in array hybridisation experiments. Nascent DNA was labelled with one fluorescent dye and G1 DNA with another. The $\log_2$ ratio of nascent over G1 DNA was calculated. For each preparation, array features that showed enrichment in nascent DNA were plotted on the UCSC human genome database (http://genome.ucsc.edu, May 2004 Assembly). The three different preparations are annotated on the left, for which blue vertical bars are positioned at the starting point of each array clone with a positive $\log_2$ ratio value. Above, regular marks of the position along chromosome 6 are shown. Below, the position of MHC regions A-M that were chosen for further analysis with real time PCR is depicted. The different regions of the MHC are highlighted in different colours: classical class I – yellow; class III – green; classical class II – blue; and extended class II – red (according to classification in Horton et al 2004).

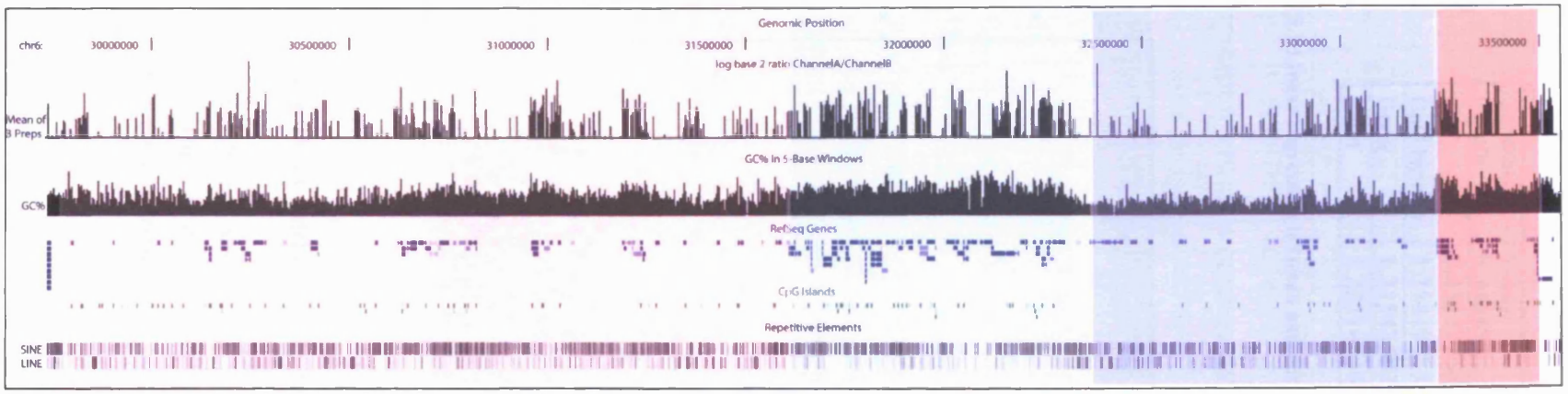


Figure 5.5 Nascent DNA abundance in the MHC region; mean values calculated from three individual preparations.

The mean value for three successful array hybridisations (one per nascent DNA preparation) was calculated for each clone. Those with enrichment in nascent DNA were plotted on the UCSC human genome database (http://genome.ucsc.edu, May 2004 Assembly). The black vertical bars on the top row are positioned at the starting point of each array clone with a positive $\log_2$ ratio value. Above, regular marks of the position along chromosome 6 are shown. Below, relevant information is depicted for the GC content, the positions of genes, CpG islands, SINE and LINE elements. The different regions of the MHC are highlighted in different colours: classical class I – yellow; class III – green; classical class II – blue; and extended class II – red (according to classification in Horton et al 2004).

Genomic feature	Y-Axis Intercept	Regression Coefficient	Correlation Coefficient
GC	0.45278754	0.016872067	0.114375832
CpG islands	1.214197157	0.008788377	0.081886726
Gene	1.215435071	0.00052596	0.025466524
Exon	1.225148543	0.006107137	0.05060425

Table 5.1 Linear regression between nascent DNA enrichment and genome statistics.

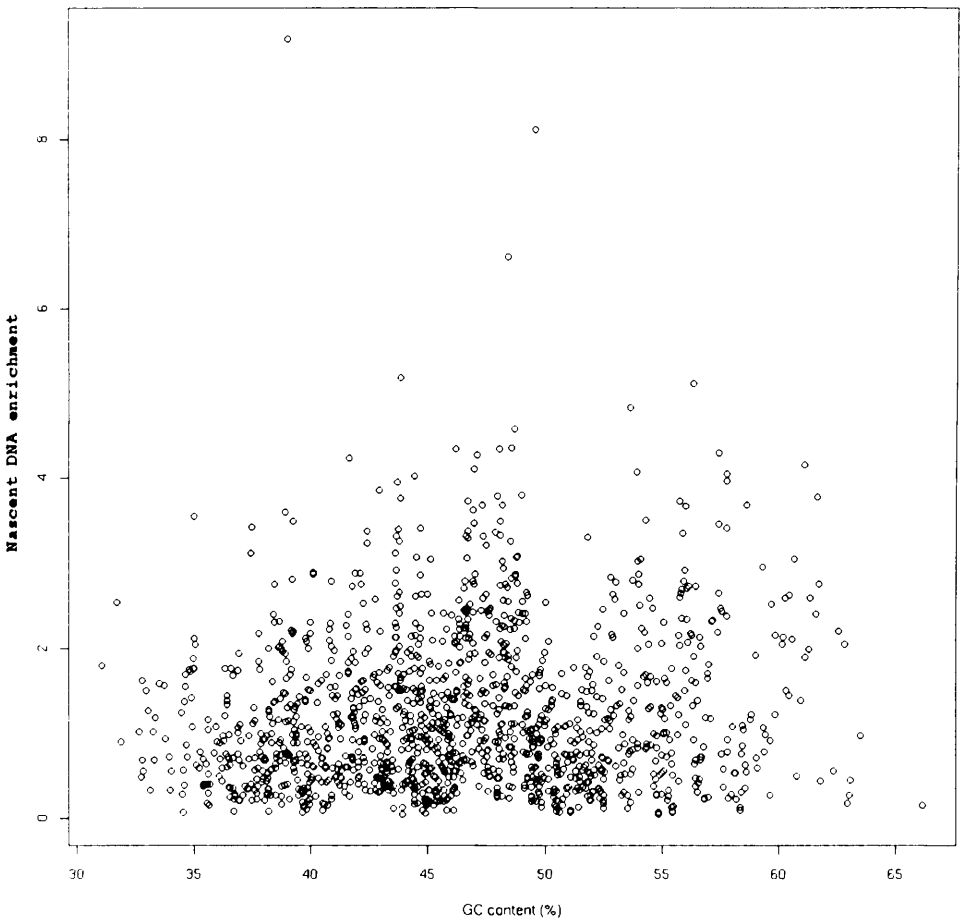


Figure 5.6 Nascent DNA enrichment in the MHC relative to GC content.

The value of GC% for each array clone is plotted against the corresponding level of nascent DNA enrichment. According to the statistical calculations, a weak positive correlation exists between nascent DNA enrichment and GC content.

However, even in cases like this, particular clones within the region were enriched to higher levels.

The mean $\log_2$ ratio for each clone across the entire region was then calculated using the results of a successful experiment from the three nascent DNA samples. Upon visual examination of the results in graphical form it appears that nascent DNA is enriched in gene-dense regions (Figure 5.5). Further analyses of the relationship between different features of the genome – GC content, or gene, exon and CpG island density – were performed relative to nascent DNA enrichment. GC content, has a weak positive correlation (> 0.1) with nascent DNA enrichment (refer to Table 5.1 and Figure 5.6).

To reduce extreme values in the mean $\log_2$ ratio dataset, two different smoothing approaches were used: the "Running mean" and "Loess" curve (Figure 5.7). The former reduces extreme values without substantially affecting the number of peaks present, whereas the latter reduces the number of peaks to a great extent.

Thirteen regions with enrichment in at least two of the three samples tested were further analysed by PCR. Four MHC class I, three MHC class III and six MHC class II sites were examined (Figure 5.4). A similar approach to that described above for the lamin B2 origin was adopted for these 13 regions.

Primer pairs spanning the regions A-M were designed at regular intervals (Figure 5.8). Standard curves were used to quantify the amount of template DNA corresponding to each primer pair. Importantly, to enable comparisons between the findings for the PGF, AHB (1) and AHB (2) samples, all the PCR reactions were prepared from a single master-mix containing each nascent DNA sample. This ensured that all reactions contained equal amount of template, provided that the volume of the master-mix used in each PCR reaction was always the same. Furthermore, the reactions were performed at the same time using the same cycling conditions. To improve accuracy and establish statistical confidence in the results, three separate amplification reactions were performed for each primer set from the same master-mix. These measures maintained uniform conditions between the different reactions for each series of experiments.

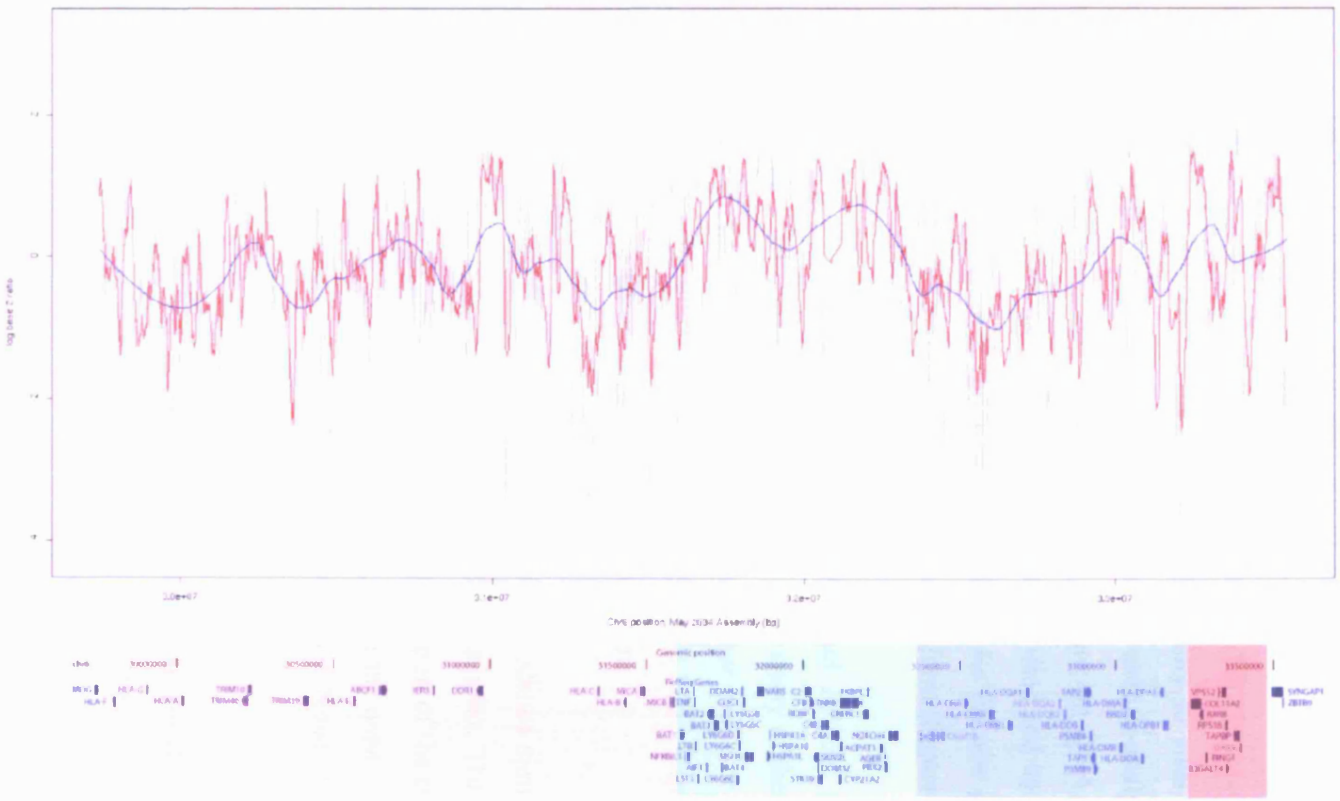


Figure 5.7 Application of smoothing approaches to log2 ratio values.

To gain a better insight into the trends of nascent DNA enrichment, reduce the levels of extreme values and minimise the amount of background signal, the array data were processed using two different smoothing approaches: "Running mean" (red line) and "Loess" (blue line). Thin grey lines represent the clones on the array. Clones of both

positive (enriched in nascent DNA) and negative (non-enriched in nascent DNA) $\log_2$ ratio value are depicted in this figure. The latter approach has a dramatic effect on the number of peaks with the parameters used here. Background highlighting as per page 49.

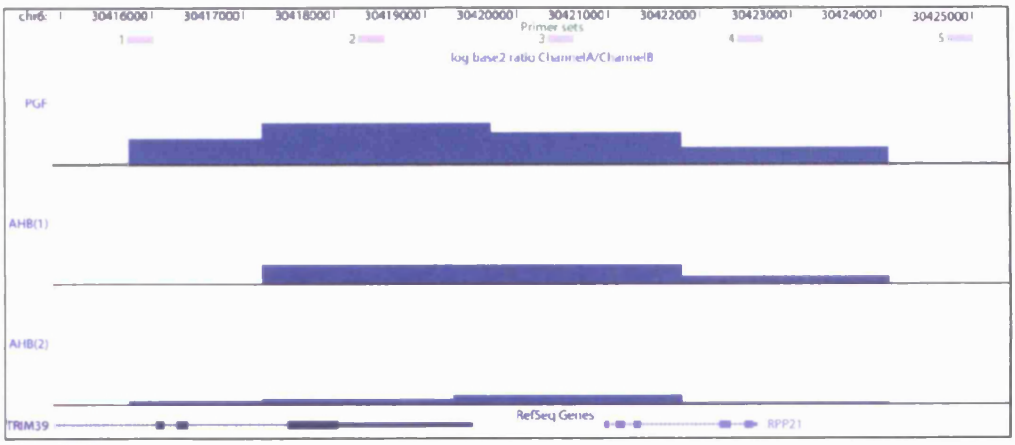
Initially, regions F, G, H and L were analysed by PCR using all three nascent DNA preparations, while the other nine regions were analysed by PGF and AHB (2) nascent DNA, as the material available for PCR reactions from AHB (1) was limited. Those four regions were selected on the basis of their proximity to sites that had been examined in Chapter 3. Subsequently, regions F, G and J were re-analysed to include some additional primer sets, using PGF and AHB (2) nascent DNA. For the latter amplification reactions more nascent DNA template was used and an annealing step of 57 °C – this temperature was found to be more suitable for the new primer sets. All this information was collated and represented in graphical form (Figure 5.8).

Region A, which overlaps with the genes *TRIM39* and *RPP21*, shows slightly different enrichment profiles of nascent DNA in the array assay of the three samples, with the greatest enrichment overlapping with either primer sets 2 or 3. PCR shows primer sets 2 and 3 to be most enriched in AHB (2) and PGF, respectively, while primer set 1 is the least enriched set in both of these samples.

Region B, which overlaps with genes *GNL1* and *PRR3,* shows similar enrichment profiles of nascent DNA in the array assay of the three samples. The most enriched clone in all three samples coincides with the promoter region of the two genes. PCR shows primer set 3, which is included in this clone, to be the most enriched in PGF while the remaining four sets are considerably less enriched. None of the five primer sets give increased enrichment in AHB (2).

Region C, which includes the *NRM* gene, shows similar enrichment profiles of nascent DNA in the array assay of PGF and AHB (1). PCR shows primer sets 1 and 4 to be the most-enriched in PGF and AHB (2), respectively, which is different from the array data.

A



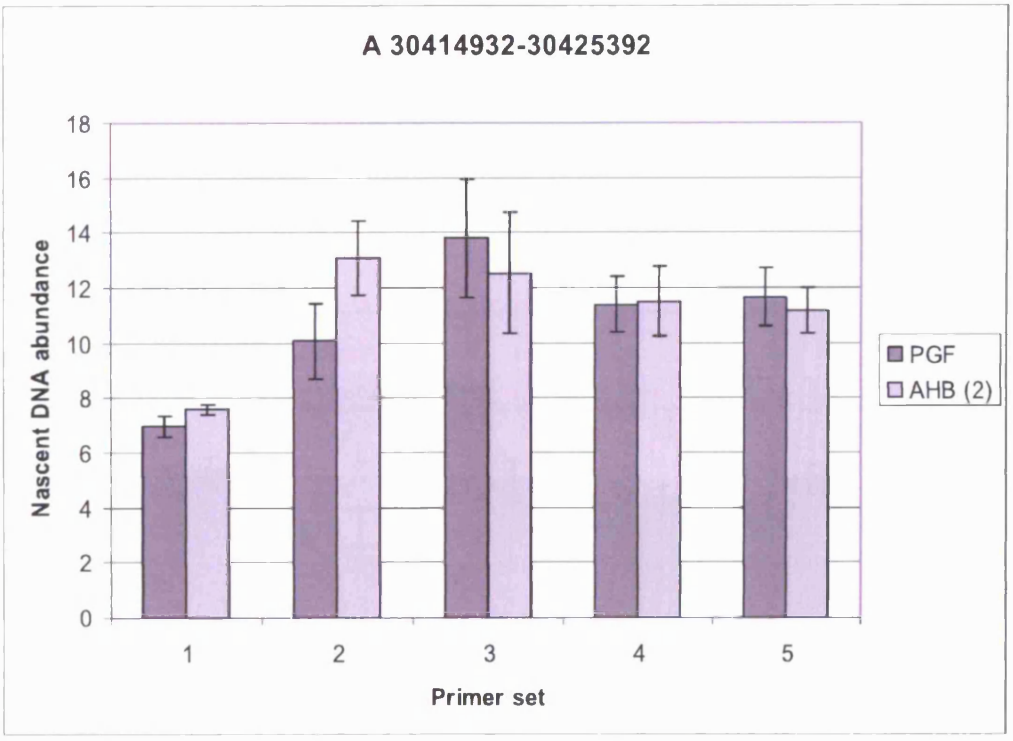
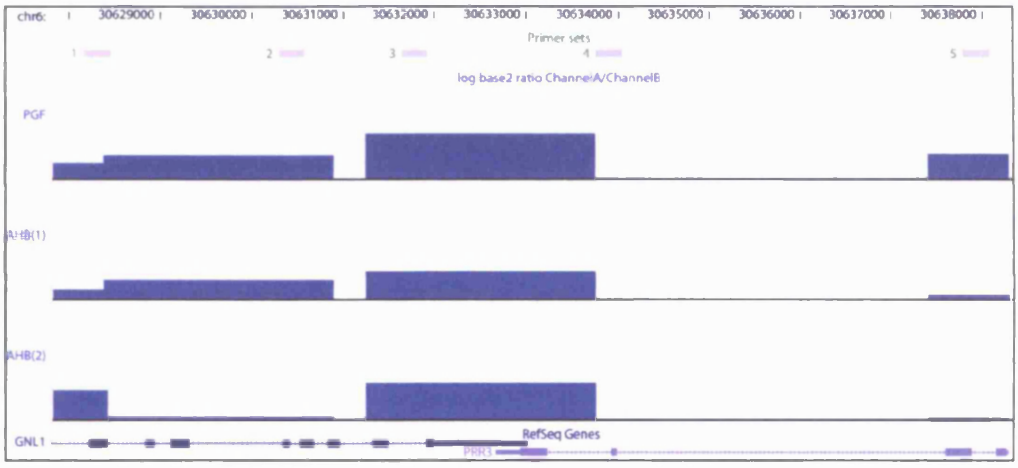


Figure 5.8 (pages 138-153) High-resolution representations of array hybridisation results and PCR analysis of nascent DNA enrichment with various primer sets in thirteen different regions (A-M) of the MHC.

From top to bottom, positions of the PCR primer sets designed in each of the thirteen regions selected (A-M) are shown (thin grey boxes represent amplicons), together with levels of enrichment for clones within the region (blue bars) and genes present (http://genome.ucsc.edu, May 2004 Assembly). Real time PCR was performed and the amount of template for each marker was calculated in accordance with a standard curve of known DNA concentrations. Same PCR conditions were used for all markers and reactions were prepared from the same master-mix of nascent DNA template to maintain uniformity within each of the two sets of PCR experiments. Bars represent the mean of three experiments with standard deviation. Further experiments would be necessary in order to allow inclusion of absent bars.

B



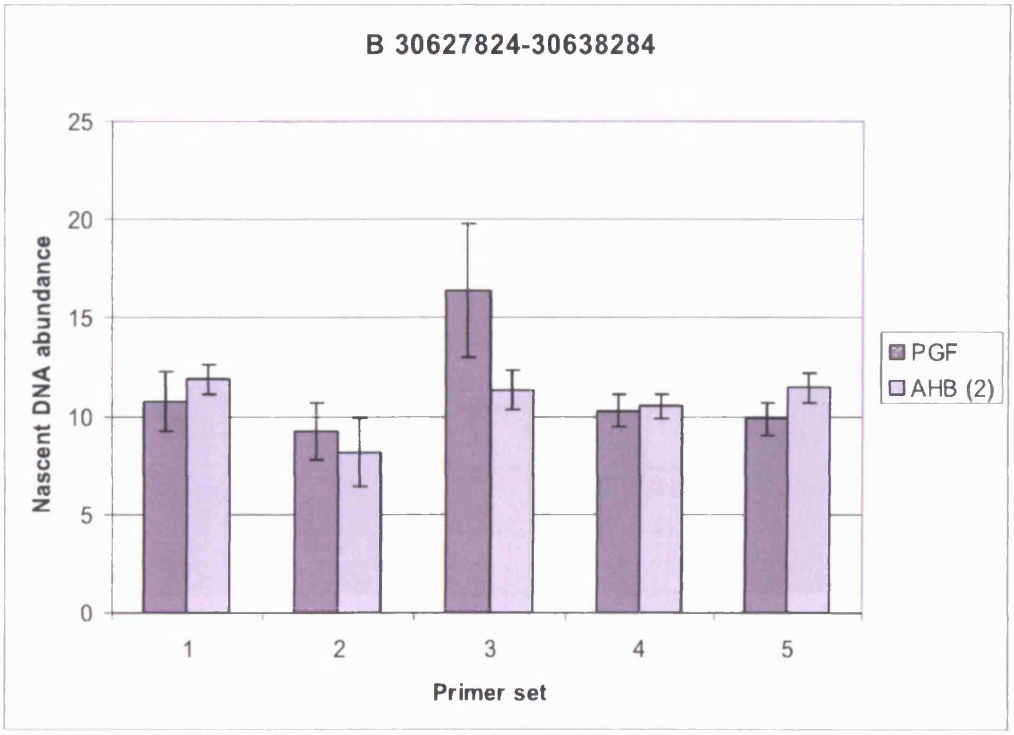
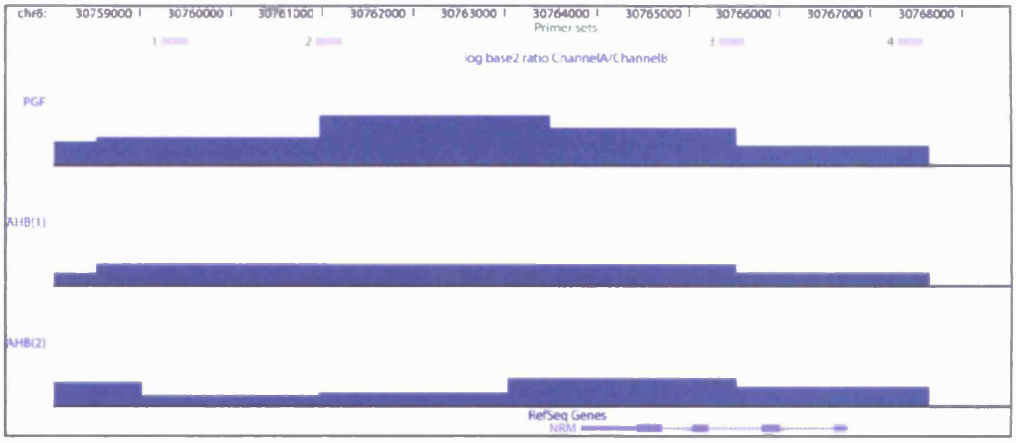


Figure 5.8 (continued) High-resolution representations of array hybridisation results and PCR analysis of nascent DNA enrichment with various primer sets in thirteen different regions (A-M) of the MHC.

Please refer to caption on page 138 for explanation.

C



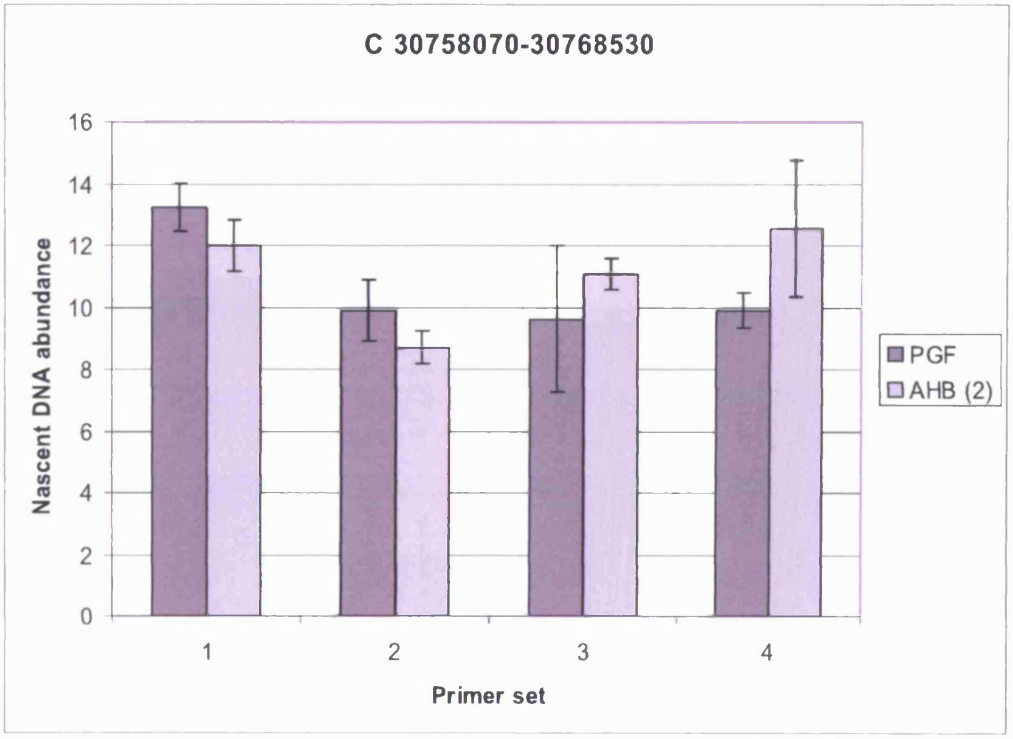
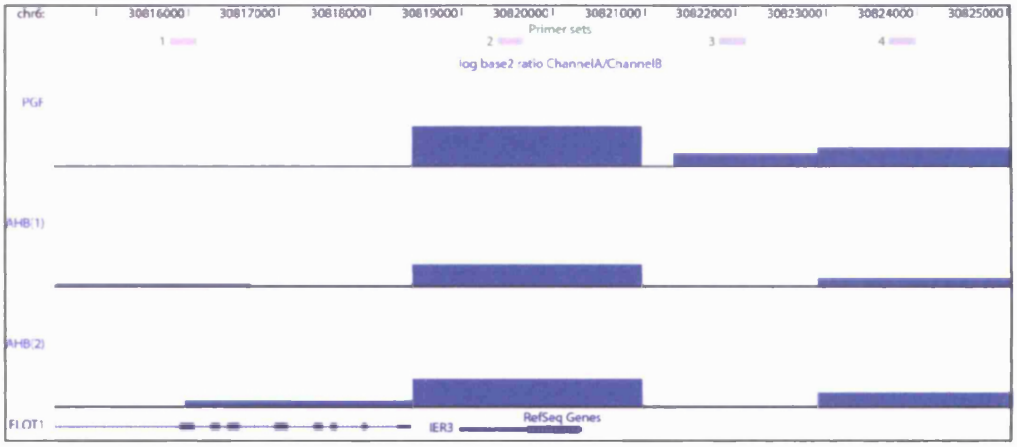


Figure 5.8 (continued) High-resolution representations of array hybridisation results and PCR analysis of nascent DNA enrichment with various primer sets in thirteen different regions (A-M) of the MHC.

Please refer to caption on page 138 for explanation.

D



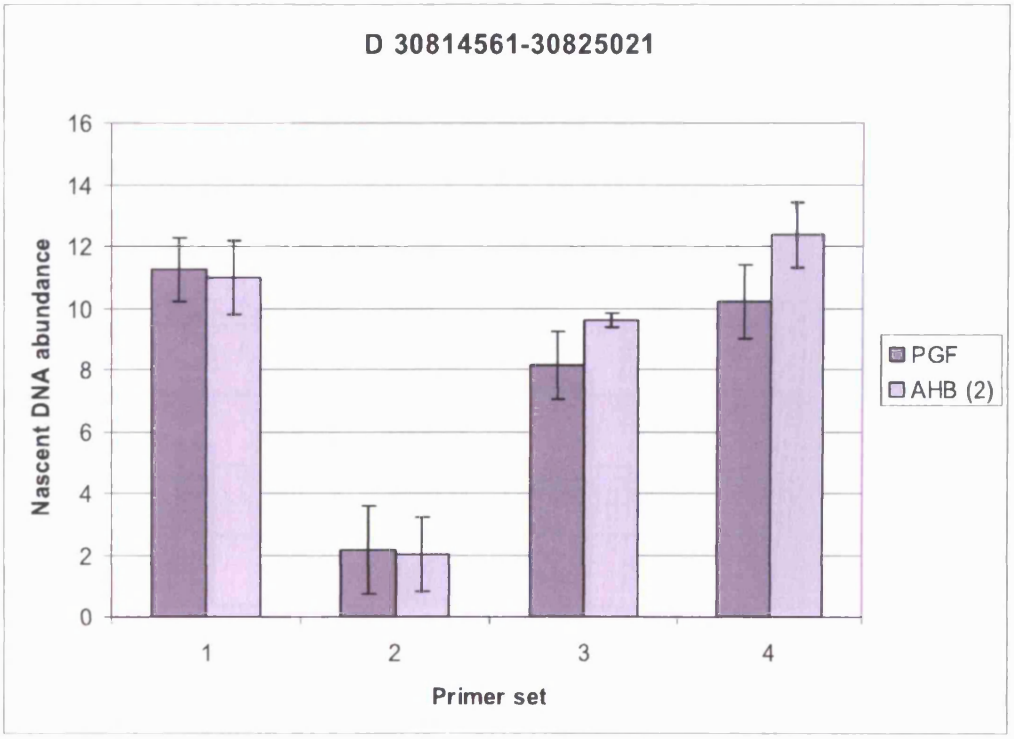
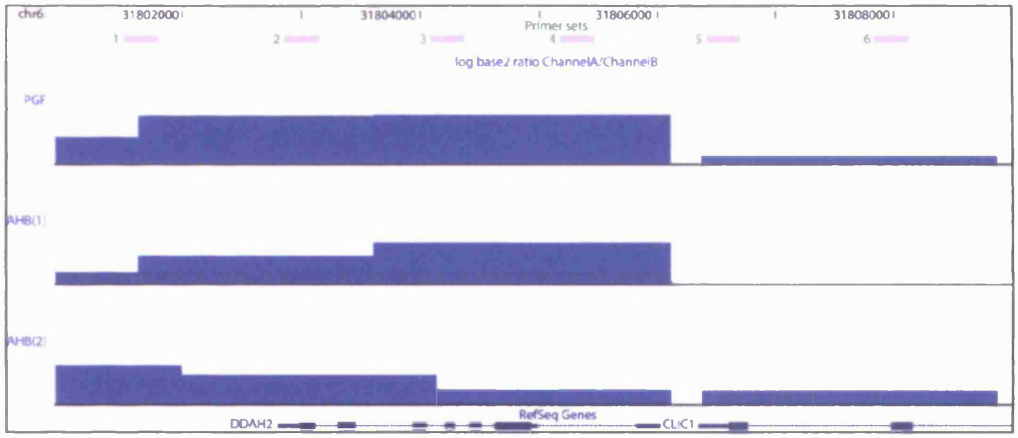


Figure 5.8 (continued) High-resolution representations of array hybridisation results and PCR analysis of nascent DNA enrichment with various primer sets in thirteen different regions (A-M) of the MHC.

Please refer to caption on page 138 for explanation.

E



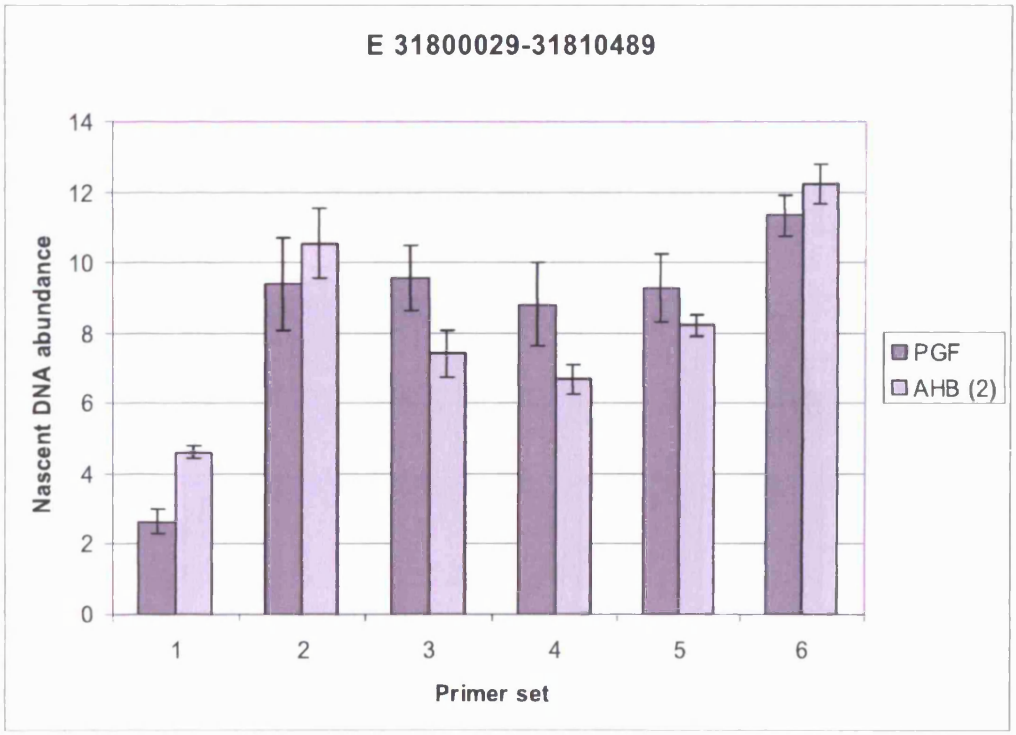
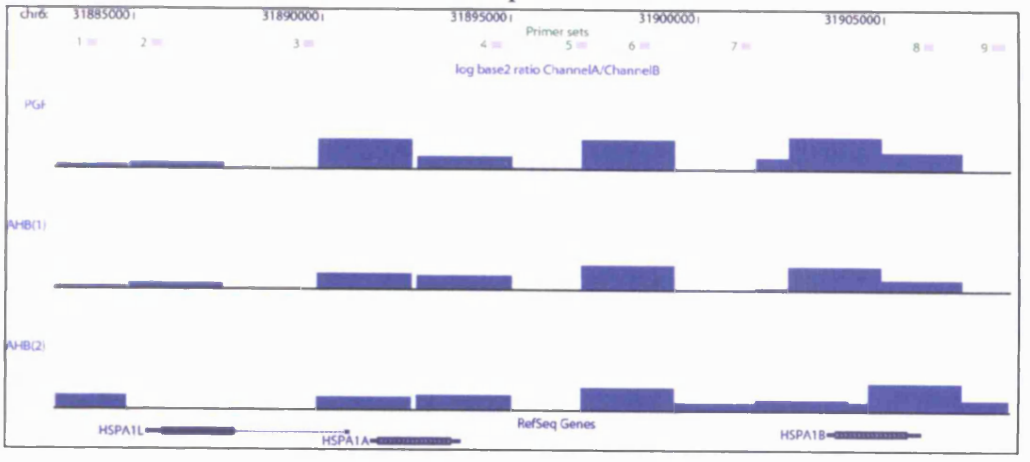
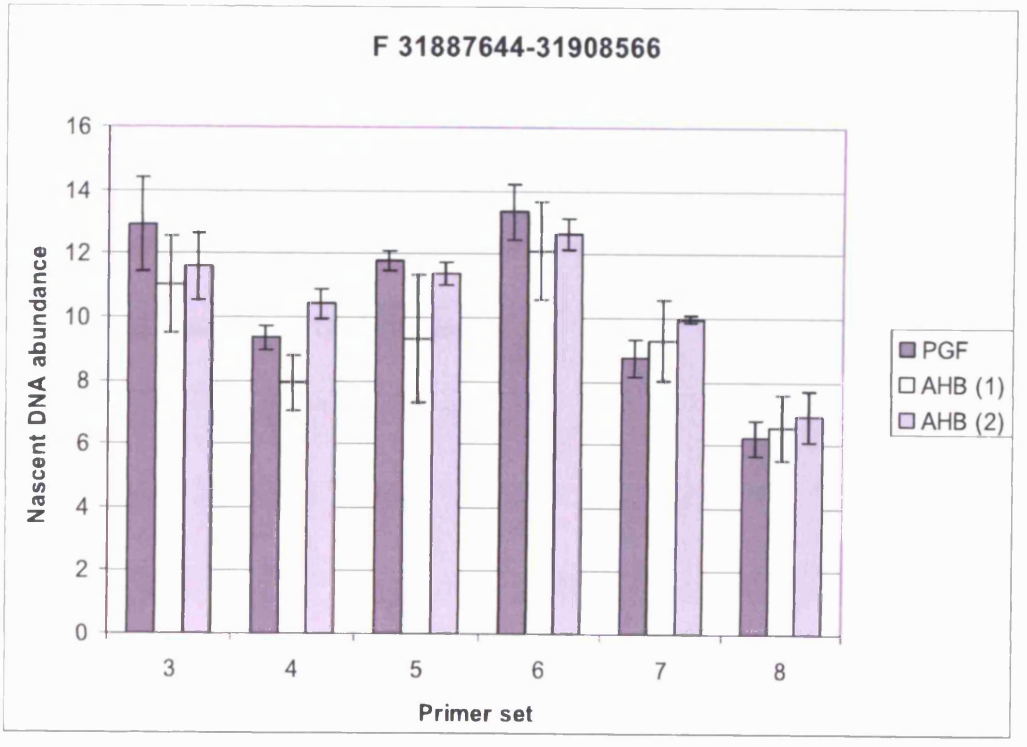


Figure 5.8 (continued) High-resolution representations of array hybridisation results and PCR analysis of nascent DNA enrichment with various primer sets in thirteen different regions (A-M) of the MHC.

Please refer to caption on page 138 for explanation.

F


chr6:
31885000
31890000
31895000
31900000
31905000
Primer sets
1
2
3
4
5
6
7
8
9
log base2 ratio ChannelA/ChannelB
PGF
AHB(1)
AHB(2)
HSPA1L
HSPA1A
RefSeq Genes
HSPA1B


F 31887644-31908566
Nascent DNA abundance
16
14
12
10
8
6
4
2
0
3
4
5
6
7
8
Primer set
PGF
AHB (1)
AHB (2)

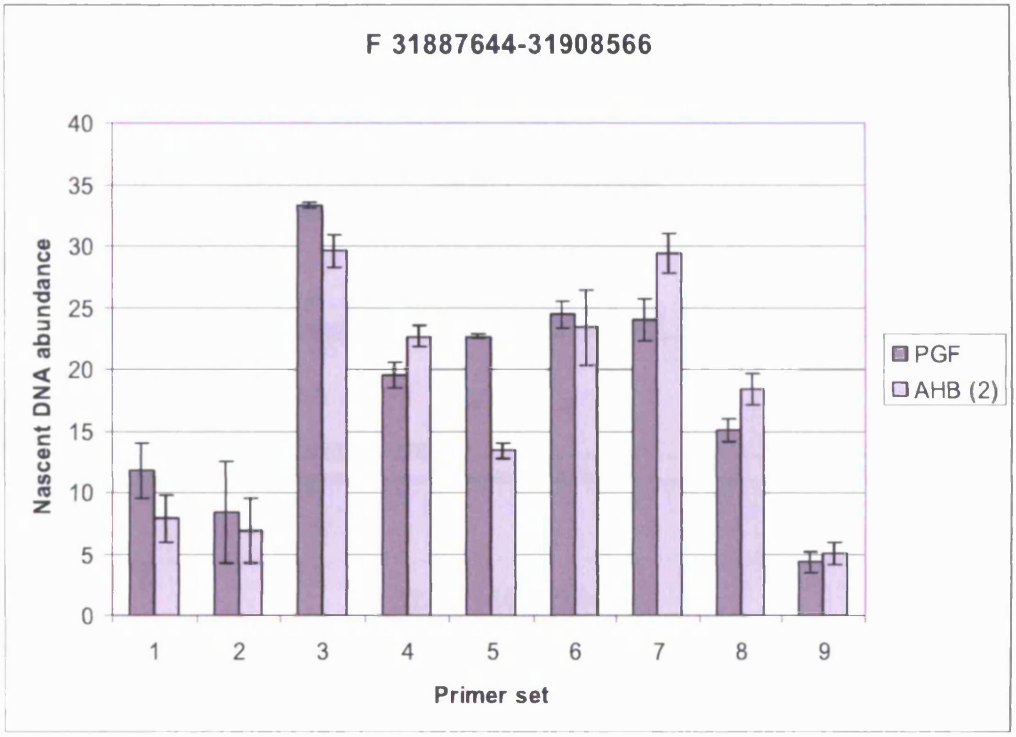
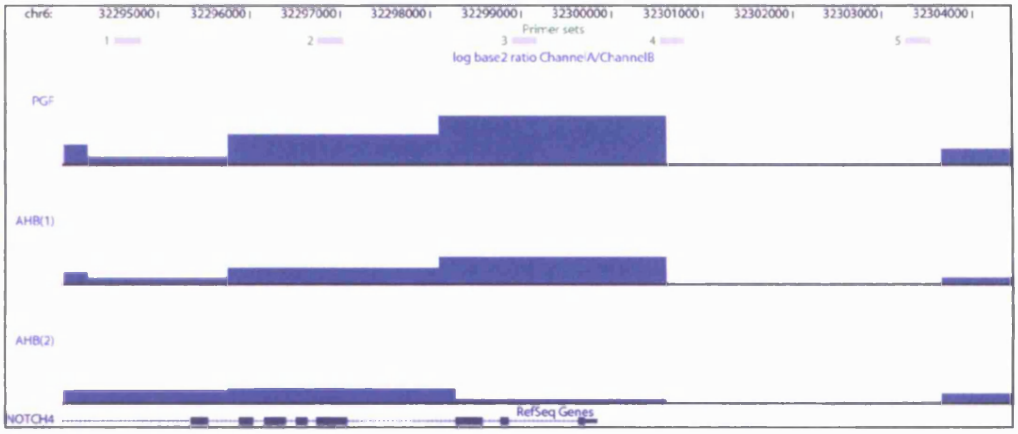
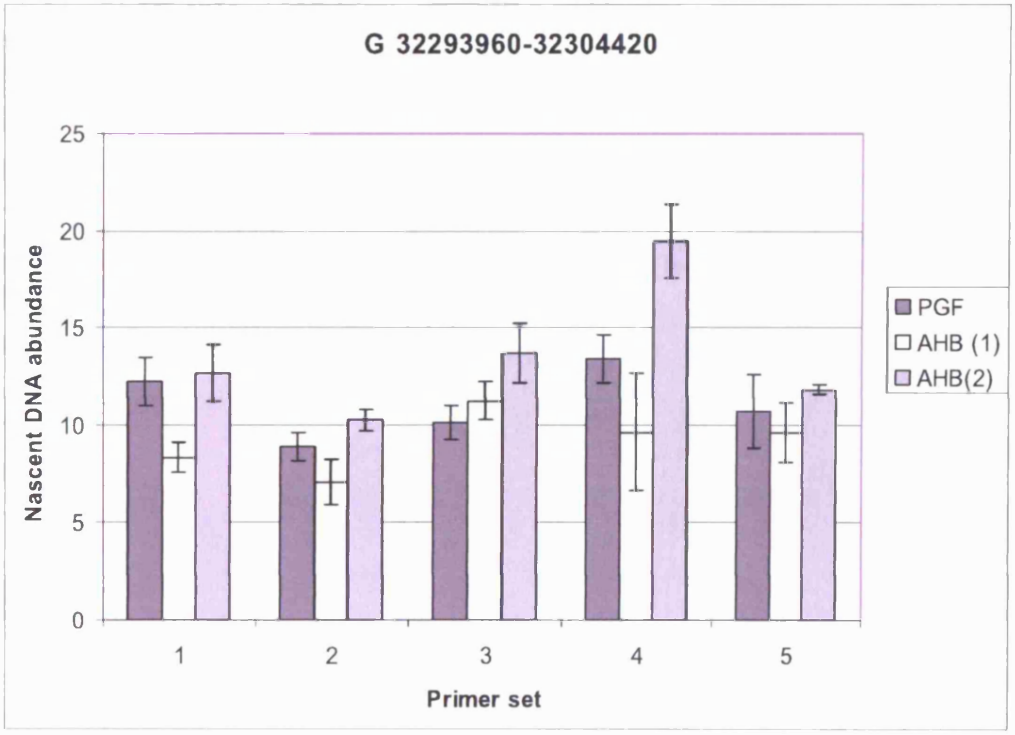


Figure 5.8 (continued) High-resolution representations of array hybridisation results and PCR analysis of nascent DNA enrichment with various primer sets in thirteen different regions (A-M) of the MHC.

Please refer to caption on page 138 for explanation.

G


chr6:
32295000
32296000
32297000
32298000
32299000
32300000
32301000
32302000
32303000
32304000
Primer sets
1
2
3
4
5
log base2 ratio Channel A/Channel B
PGF
AHB(1)
AHB(2)
NOTCH4
RefSeq Genes


G 32293960-32304420
Nascent DNA abundance
25
20
15
10
5
0
1
2
3
4
5
Primer set
PGF
AHB (1)
AHB(2)

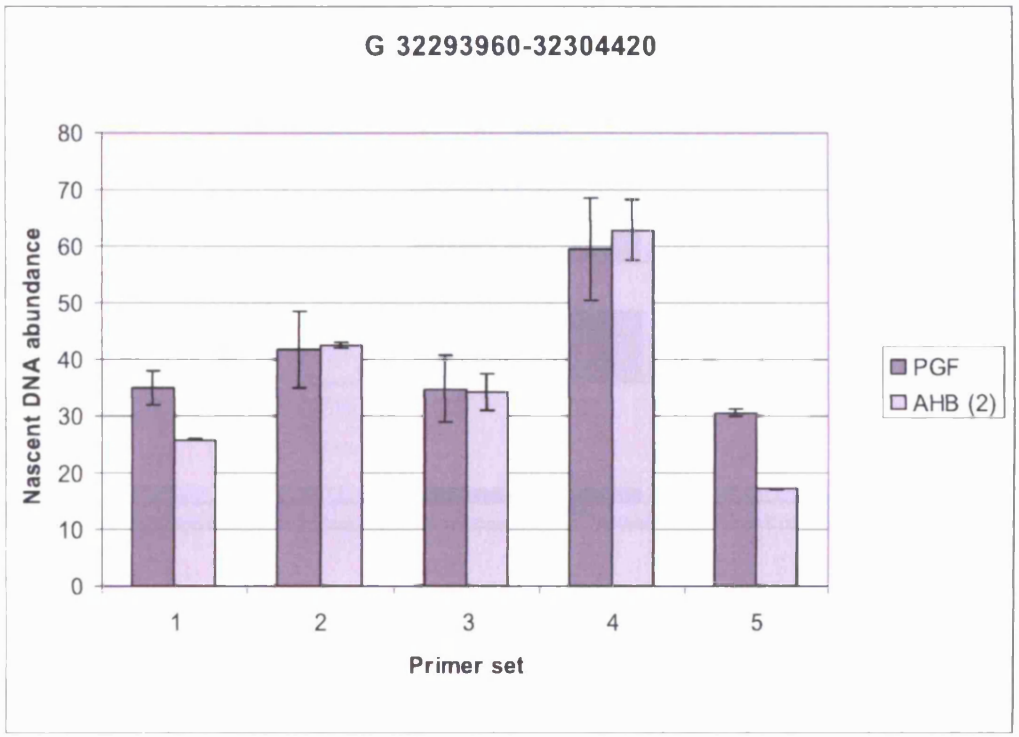
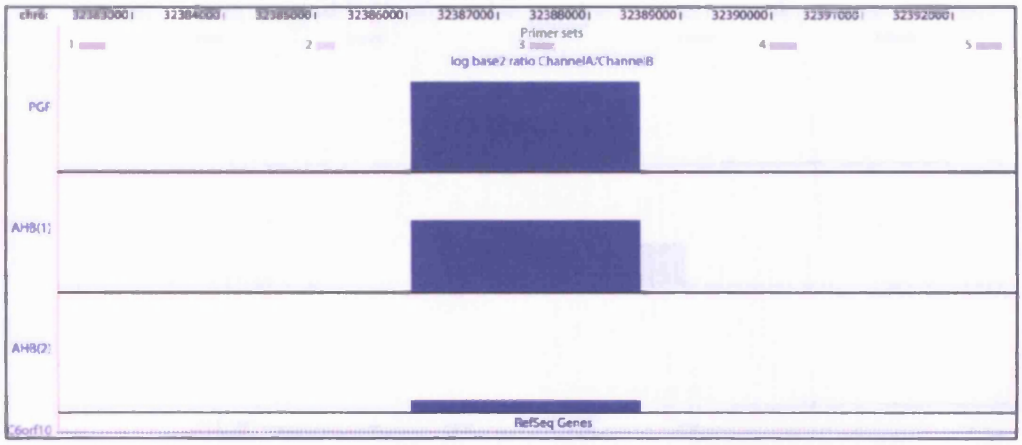


Figure 5.8 (continued) High-resolution representations of array hybridisation results and PCR analysis of nascent DNA enrichment with various primer sets in thirteen different regions (A-M) of the MHC.

Please refer to caption on page 138 for explanation.

H



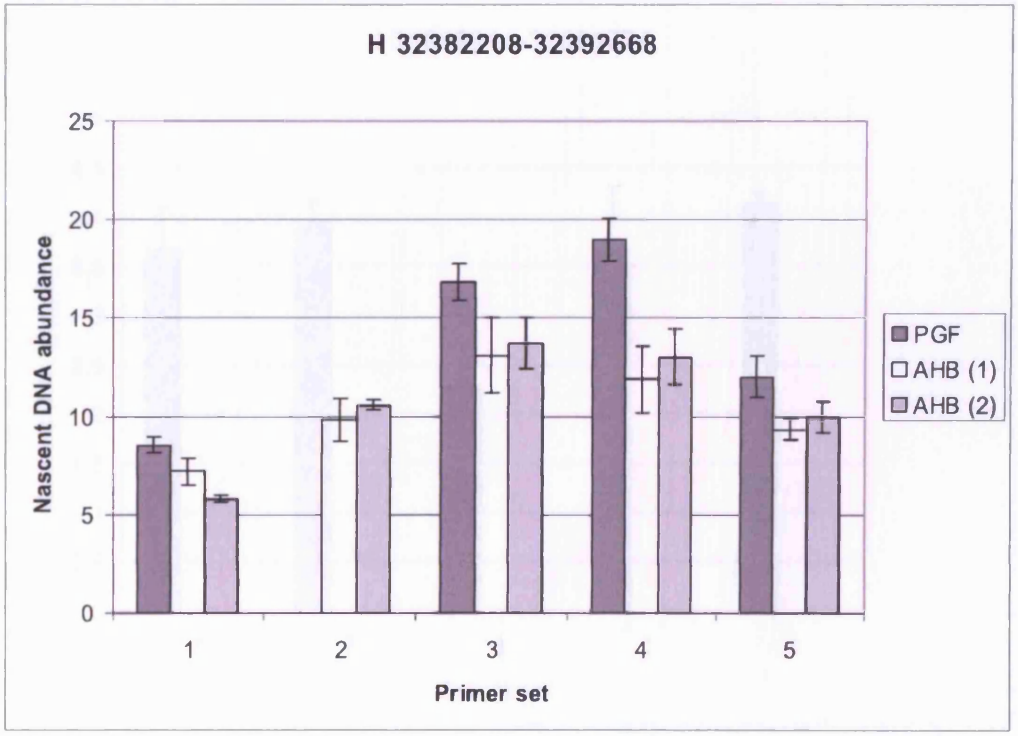
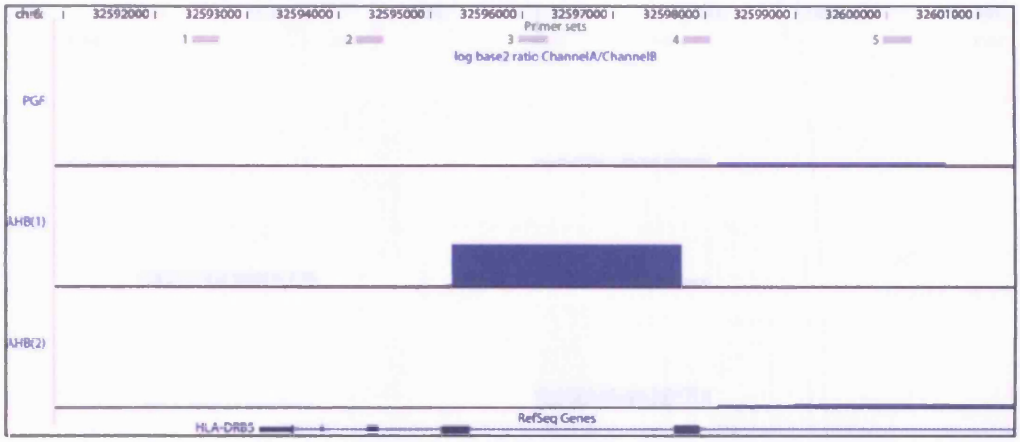


Figure 5.8 (continued) High-resolution representations of array hybridisation results and PCR analysis of nascent DNA enrichment with various primer sets in thirteen different regions (A-M) of the MHC.

Please refer to caption on page 138 for explanation.

I



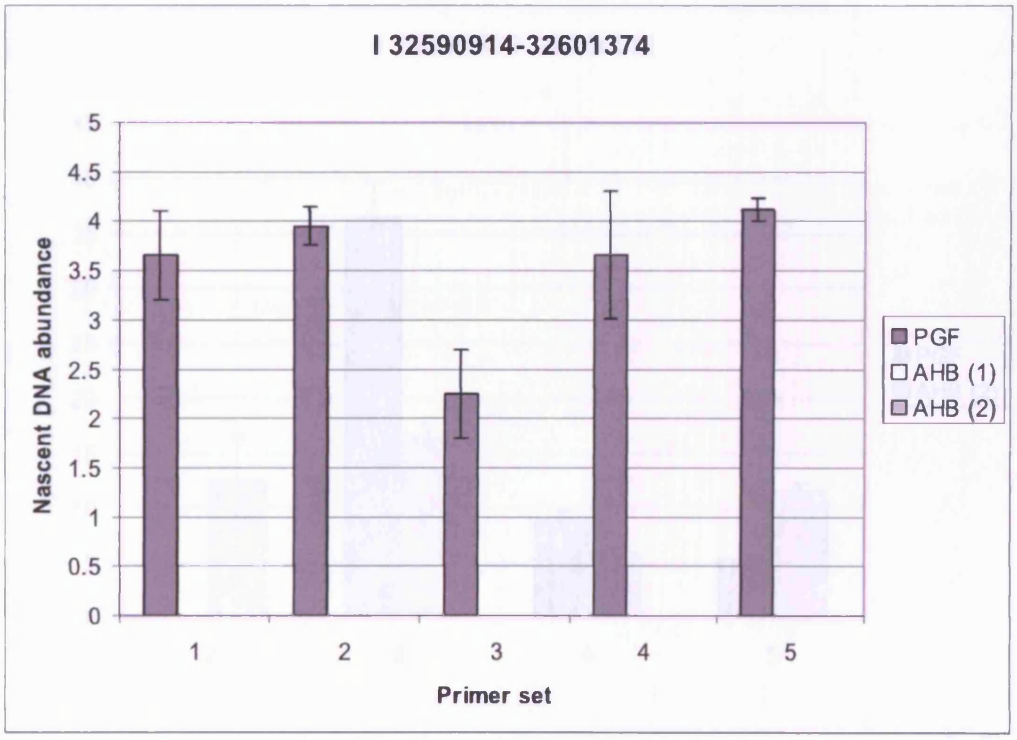
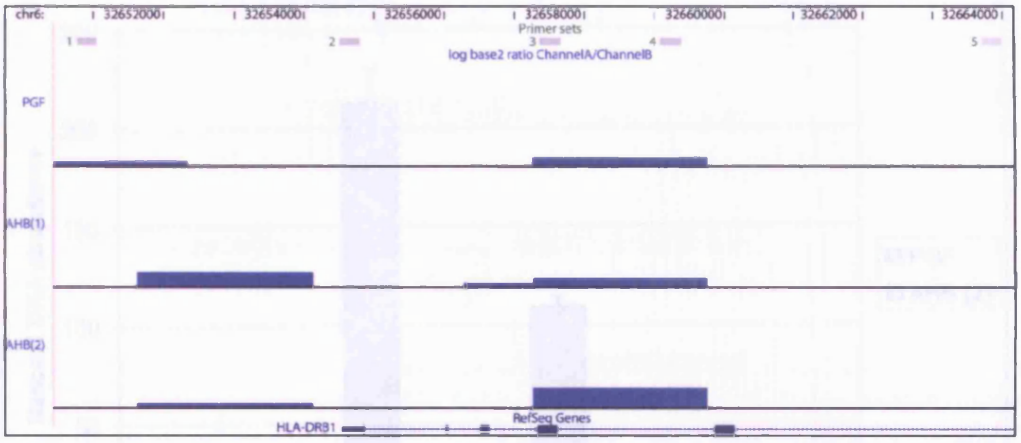
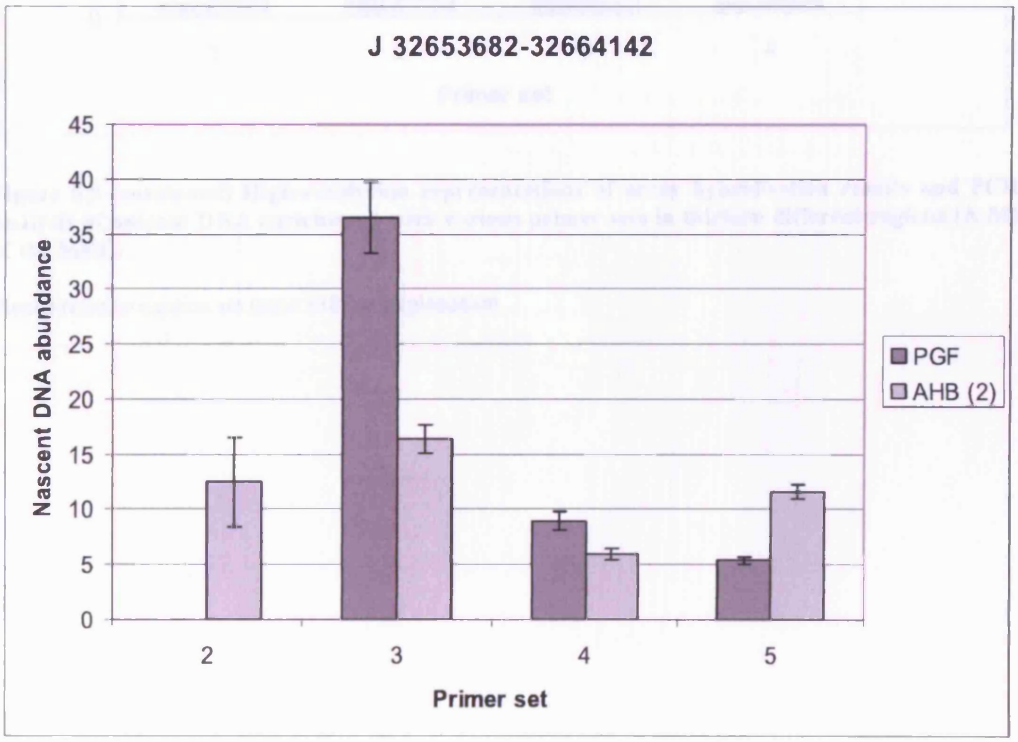


Figure 5.8 (continued) High-resolution representations of array hybridisation results and PCR analysis of nascent DNA enrichment with various primer sets in thirteen different regions (A-M) of the MHC.

Please refer to caption on page 138 for explanation.

J


chr6:
32652000
32654000
32656000
32658000
32660000
32662000
32664000
Primer sets
1
2
3
4
5
log base2 ratio ChannelA/ChannelB
PGF
AHB(1)
AHB(2)
RefSeq Genes
HLA-DRB1


J 32653682-32664142
Nascent DNA abundance
45
40
35
30
25
20
15
10
5
0
2
3
4
5
Primer set
PGF
AHB (2)

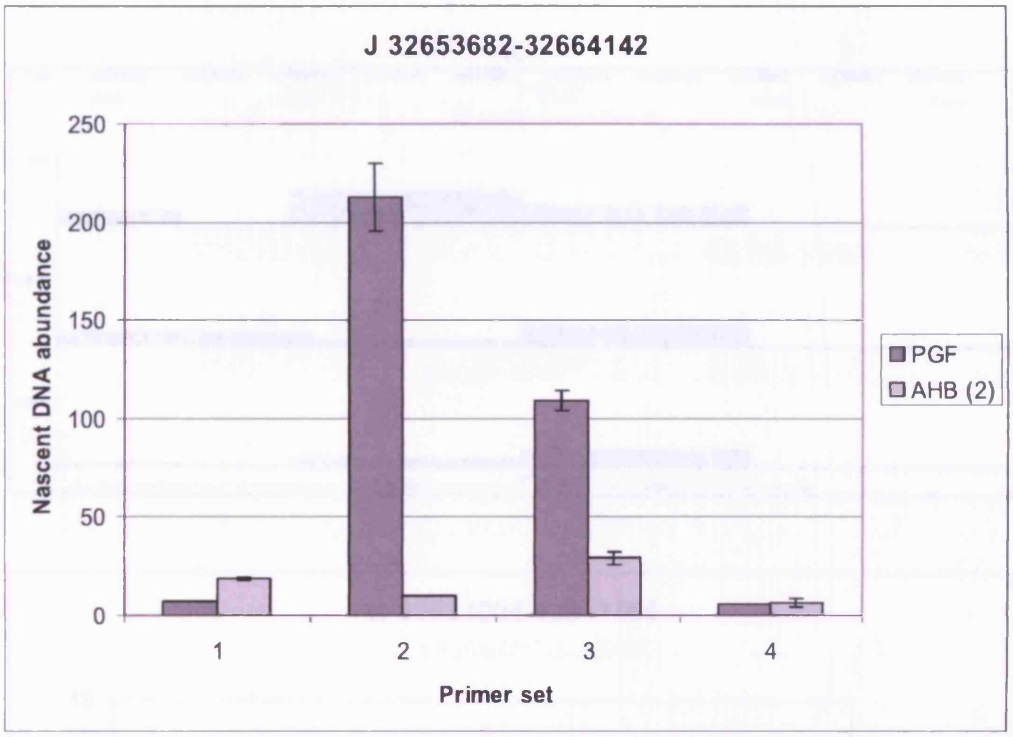
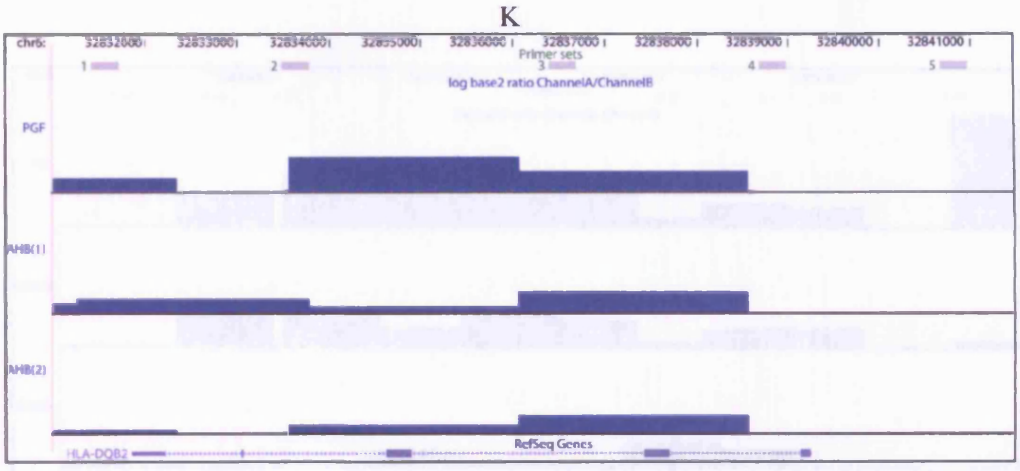


Figure 5.8 (continued) High-resolution representations of array hybridisation results and PCR analysis of nascent DNA enrichment with various primer sets in thirteen different regions (A-M) of the MHC.

Please refer to caption on page 138 for explanation.



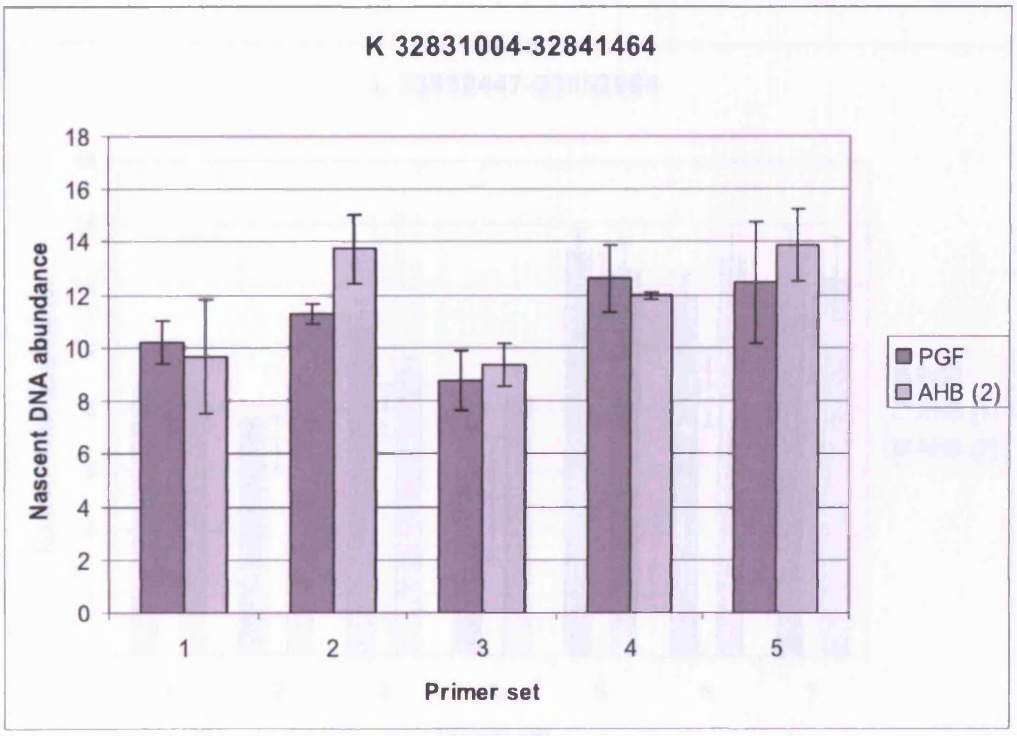
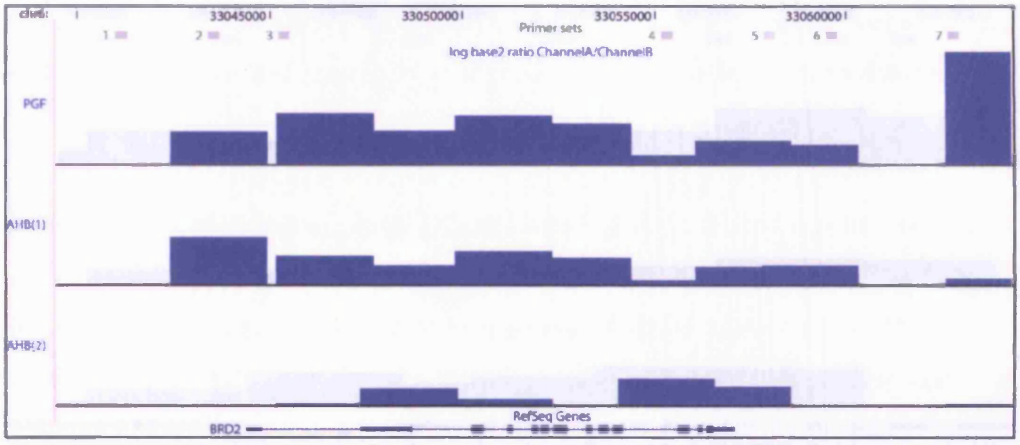


Figure 5.8 (continued) High-resolution representations of array hybridisation results and PCR analysis of nascent DNA enrichment with various primer sets in thirteen different regions (A-M) of the MHC.

Please refer to caption on page 138 for explanation.

L



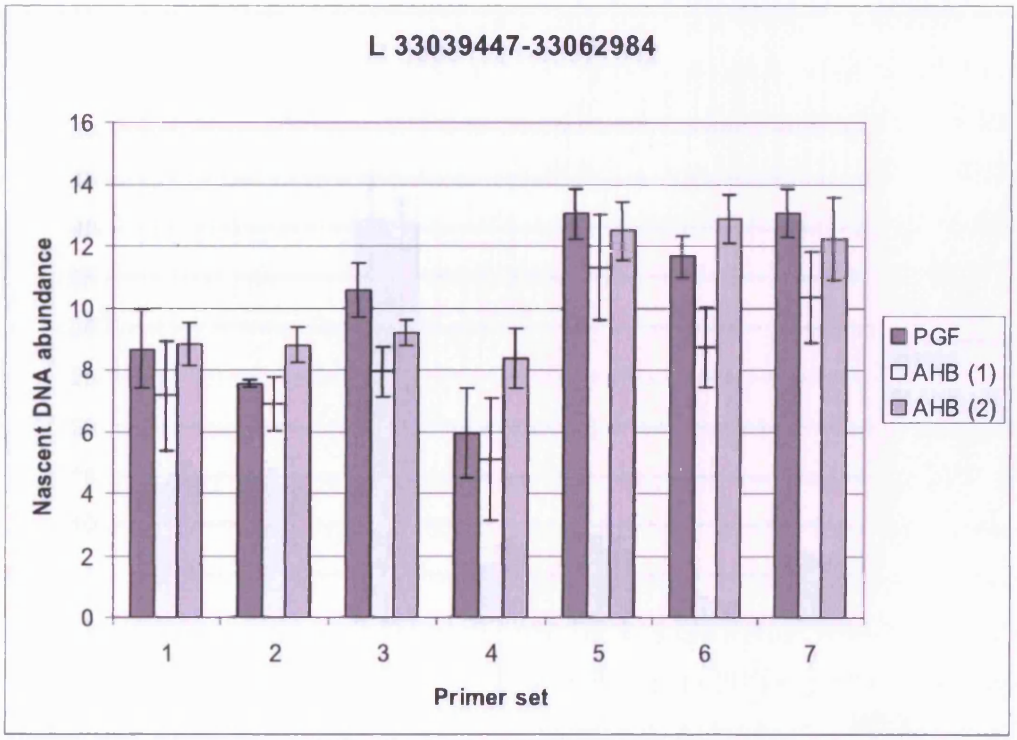
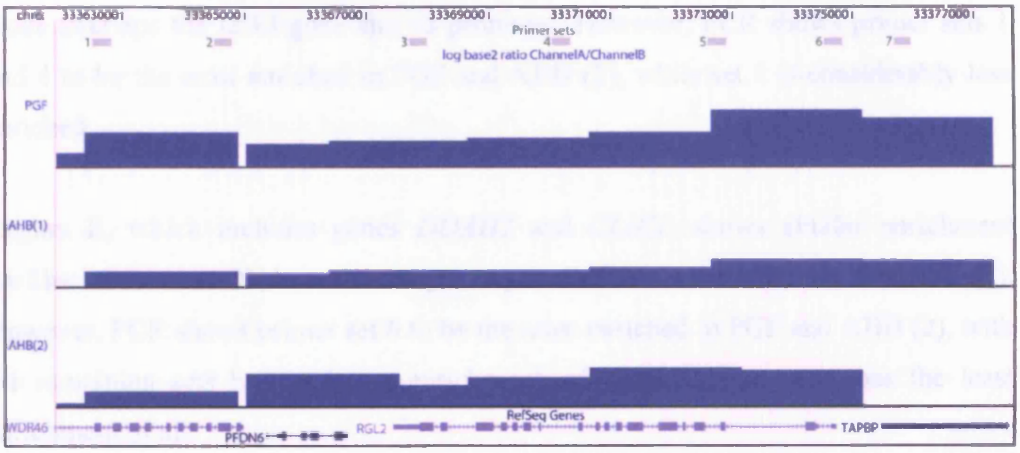


Figure 5.8 (continued) High-resolution representations of array hybridisation results and PCR analysis of nascent DNA enrichment with various primer sets in thirteen different regions (A-M) of the MHC.

Please refer to caption on page 138 for explanation.

M



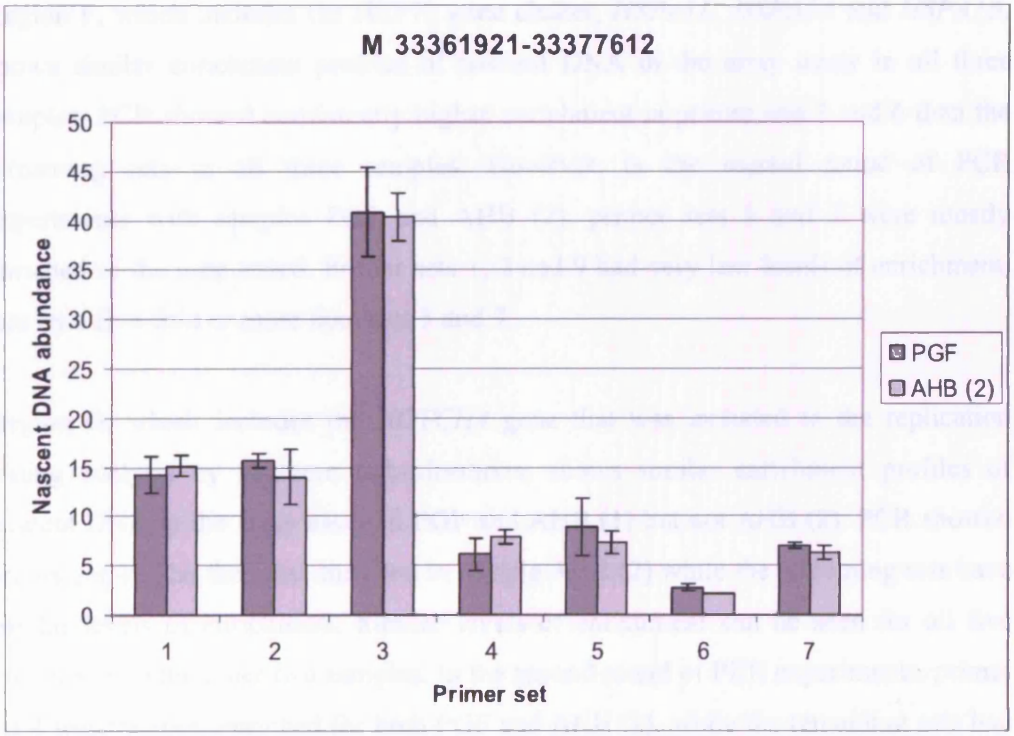


Figure 5.8 (continued) High-resolution representations of array hybridisation results and PCR analysis of nascent DNA enrichment with various primer sets in thirteen different regions (A-M) of the MHC.

Please refer to caption on page 138 for explanation.

Region D, which includes the *FLOT1* and *IER3* genes, shows similar enrichment profiles of nascent DNA in the array assay of all three samples. The most enriched clone overlaps the *IER3* gene and its promoter. However, PCR shows primer sets 1 and 4 to be the most enriched in PGF and AHB (2), while set 2 is considerably less enriched.

Region E, which includes genes *DDAH2* and *CLIC1*, shows similar enrichment profiles of nascent DNA in the array assay in PGF and AHB (1) but not AHB (2). However, PCR shows primer set 6 to be the most enriched in PGF and AHB (2), with the remaining sets having lower enrichment, of which primer set 1 has the least enrichment of all.

Region F, which includes the HSP70 gene cluster, *HSPA1L*, *HSPA1A* and *HSPA1B,* shows similar enrichment profiles of nascent DNA in the array assay in all three samples. PCR showed consistently higher enrichment in primer sets 3 and 6 than the remaining sets in all three samples. However, in the second round of PCR experiments with samples PGF and AHB (2), primer sets 3 and 7 were mostly enriched of the nine tested. Primer sets 1, 2 and 9 had very low levels of enrichment, less by a five-fold or more than sets 3 and 7.

Region G, which includes the *NOTCH4* gene that was included in the replication timing analysis by southern hybridisations, shows similar enrichment profiles of nascent DNA in the array assay in PGF and AHB (1) but not AHB (2). PCR showed primer set 4 to be the most enriched in sample AHB (2) while the remaining sets have similar levels of enrichment. Similar levels of enrichment can be seen for all five primer sets in the other two samples. In the second round of PCR experiments, primer set 4 was the most enriched for both PGF and AHB (2), while the remaining sets had lower levels of enrichment.

Region H, which overlaps with *C6orf10* that was included in the replication timing analysis of the previous chapter, has considerable similarity in the array results of the three samples. Primer sets 3 and 4 are the most enriched of those examined with the three samples, primer set 4 being the most enriched for sample PGF while primer set 3 is most enriched for samples AHB (1) and AHB (2).

Region I, which covers the *HLA-DRB5* gene, does not show any consistency between the array data from the three samples. It has very low enrichment for the five primer sets examined with sample PGF, while samples AHB (1) and AHB (2) have no enrichment for any of these primer sets.

Region J, which covers *HLA-DRB1*, has more similarity in the array findings between the two AHB samples than with PGF. In the first round of PCR experiments it is highly enriched at primer set 3 for nascent DNA sample PGF. For AHB (2) the same primer set is also the most enriched, but not at the same level. In the second round of PCR experiments, primer set 2 is the most enriched for sample PGF. Overall, there are much lower levels of enrichment for sample AHB (2) than for PGF.

Region K, which overlaps with *HLA-DQB2*, has limited similarity in the array results of the two AHB samples. It has similar levels of enrichment across the five primer sets examined. For both samples PGF and AHB (2), primer set 5 has the highest enrichment.

Region L, which includes the gene *BRD2*, shows similarities in the array findings between PGF and AHB (1). Primer sets 5, 6 and 7 have similar high levels of enrichment for all three samples.

Finally, region M, which overlaps four different genes including *TAPBP*, has a high level of similarity between PGF and AHB (1). This region has a very high enrichment with primer set 3 for both PGF and AHB (2), while the remaining sets have considerably lower levels of enrichment. This is the only example of a region where for both nascent DNA samples one primer set is remarkably more enriched than the rest, twice as much or higher. Nevertheless, the array data are generally not in agreement with the findings of the PCR experiments for this region.

5.3 Discussion

Nascent DNA was isolated from PGF and AHB lymphoblastoid cells. The material from three different biological samples was then used to perform hybridisations on the

MHC tiling path arrays. Examination of the graphical representations of the results indicates that there is a high level of similarity between the three sets of data. Nevertheless, at higher resolution, samples PGF and AHB (1) share more similarities than samples AHB (1) and AHB (2), which originate from the same cell line (figure 5.5). Thirteen regions were chosen for further examination with several PCR primers each. Overall, there is congruence in the results of the nascent DNA samples. Evidently, for the second set of PCR reactions there is a higher level of enrichment, which is likely due to the increased amount of DNA template used per reaction. The statistical analysis performed on the synoptic array data shows a weak positive correlation between GC content and nascent DNA enrichment. This study identifies sites that likely correspond to origins of replication, and provides information about the frequency and distribution of replication initiation sites within the MHC. It also offers an insight to the number of origins present across this region.

Despite the similarity between the findings for the three samples, the same conclusion is not always true with respect to the findings of the two methods. For example, as in the case of region D, the clone overlapping *IER3* is the most enriched in that region, while primer set 2 which covers part of that gene is the least enriched of the four. Analysis of PCR experiments showed that PCR amplification with primer sets overlapping CpG islands, such as primer set 2 in region D[1], was difficult for the annealing temperatures used. It is a significant shortcoming for this PCR assay as some reports have suggested that replication origins localise at or near CpG islands (Vassilev and Johnson 1990; Taira, Iguchi-Ariga et al. 1994; Paixao, Colaluca et al. 2004).

Although the amount of template present for each primer set was determined in accordance with a standard curve, the precision of those standard curves varied from one set to another. Since the efficiency of each reaction is different, as is the range of DNA concentrations within the linear dynamic range of the reaction, template quantitation with this approach is not entirely flawless.

[1] http://genome.ucsc.edu/cgi-bin/hgTracks?hgsid=84744736&db=hg17&position=chr6%3A30818287-30820419 Follow link for precise localisation of the CpG island mentioned in the text body.

In order to attain good understanding of the PCR data it is important to consider the basic principles of the method. The reaction is influenced by the PCR cycling conditions; a particular temperature will be optimal for certain primer pairs to anneal but less fitting for others. Furthermore, primer concentration could influence the kinetics of the reaction as would the amount of template present. Special attention was given to the size of the PCR products in an attempt to maintain uniformity in this respect. Nevertheless, the amplicon produced by each set of primers is unique and has unique characteristics. Generally, it was difficult to avoid small variations in amplicon size as well as differences in GC% composition. This heterogeneous amplicon structure could also influence the progress of each reaction. Ideally the template and primer concentration as well as the cycling conditions for each reaction should be optimised experimentally. However, such a measure, if taken, would upset the uniformity of conditions between the different reactions, introducing numerous variables.

The array experiments have an important advantage over the PCR approach. They give a global overview of nascent DNA enrichment as opposed to the local focus the PCR strategy provides. In turn, this offers a better insight if comparisons need to be made between different sections within the MHC. However, both experimental approaches presented an unexpected phenomenon.

The array hybridisation experiments show several portions of the MHC, in some cases spanning ~100 Kb, which consist of clones that are enriched in nascent DNA. What is the underlying reason for this observation? Should nascent DNA not indicate enrichment in 2-3 neighbouring clones only?

As many as seven hundred MHC clones indicate nascent DNA enrichment. Although some of these clones are overlapping, the existence of such a large number of replication initiation sites, one every ~7 Kb, does not comply with the proposed frequency in the human genome of one origin in approximately every 100 Kb (Huberman and Riggs 1968). Therefore, it is a reasonable assumption that several of the positive features on the array contain background signal. This is not to be confused with the background signal that is detected throughout the arrays after scanning but is subtracted from the foreground signal as part of the normalisation

process. The background signal in question refers to clones that may have acquired their signal as a result of non-specific, non-nascent DNA that has gone through the isolation process, for example, during fraction collection. To prepare the hybridisation mix with equal amounts of nascent and G1 DNA, the concentration of the former was calculated according to the absorbance of single-stranded nucleic acid molecules (1 A_{260} ssDNA = 37 μg/ml). It is possible, however, that some of the molecules in solution were double stranded, thus introducing errors in the calculation and subsequent steps. The background signal witnessed is a likely consequence of such an error too. The normalisation methods employed did not remove such features from the data sample. Subsequently, the normalised data were reduced using data smoothing approaches and application of empirically defined parameters in order to converge towards the proposed models. The "Running mean" plot appears to be more reliable than the "Loess" curve, based on the parameters that were employed here.

Background signal in the results appears also to exist in the findings of the PCR experiments. For most primer sets examined some level of enrichment was present. In principle enrichment should only be seen in those primers that fulfill the nascent DNA selection criteria. Therefore, it is likely that the nearly-horizontal level of enrichment detected in most regions corresponds to background. For the first set of PCR experiments the background level is around 10 on the scale of nascent DNA abundance. For the second set of experiments the background level is higher, twice or three times that of the early experiments. Nevertheless, it is also possible that some of the primer sets represented by lower peaks, just above the perceived background level, are close to the initiation sites but do not exactly overlap. The most enriched sets, on the other hand, probably correspond to initiation sites.

The PCR reactions across region I, which overlaps the *HLA-DRB5* gene, indicated very low levels of enrichment for the nascent DNA sample PGF, while no enrichment was observed for the two AHB samples. According to previous studies (Dayan, Londei et al. 1991; Stewart, Horton et al. 2004), the MHC haplotypes for these two cell lines are quite different. Haplotype is the term used to describe the genetic constitution of a specific genomic region. With respect to variation in the organisation of the *HLA-DRB* genes, AHB is homozygous for the DR3 haplotype, which contains *HLA-DRB1*, *-DRB2*, *-DRB3* and *-DRB9*. PGF on the other hand is homozygous for a

different haplotype, DR15, which contains *HLA-DRB1*, *-DRB6*, *-DRB5* and *-DRB9*. The absence of the *HLA-DRB5* gene from AHB cells explains the lack of enrichment for the five primer sets tested. The low levels of enrichment detected across region I likely amount to background signal.

Region J is another case that could be justified by the haplotype differences mentioned above as the PCR experiments across the region show a substantial variation between the two nascent DNA samples. Two explanations can be offered on the basis of haplotype difference: (i) that there is substantial sequence heterogeneity between the two cell lines for those particular primer sets, consequently affecting the efficiency of the PCR reaction; (ii) that an origin of replication is active in PGF but not AHB cells. The discrepancy observed between the first and second round of experiments with respect to primer set 2 needs further experiments in order to be clarified.

The arrays do not contain any well characterised origin sequences that could be used as controls for these experiments. This was the case despite the fact that an origin of replication had been characterised in the HSP70 cluster, in the MHC class III region. The experiments by Taira *et al.* (Taira, Iguchi-Ariga et al. 1994) involved PCR amplification of nascent DNA with primers pairs that had been designed based on sequence information that was available at the time. However, precise mapping of those primer sequences is not possible with reference to a recent release of the human genome (UCSC Genome Browser - May 2004 Assembly). Similar sequences can be identified, some of which are represented more than once within the HSP70 cluster. This observation is in agreement with the fact that the HSP70 cluster contains several regions with high levels of homology at or near the three genes. A relatively low annealing temperature of 52 °C was used for the PCR reactions in those experiments, possibly allowing the formation of non-specific amplification products.

Briefly, in Taira *et al.* (Taira, Iguchi-Ariga et al. 1994), different centrifugation fractions were used to isolate nascent DNA with different sizes and then PCR amplification was performed. The PCR products were blotted on a membrane and then probed with radioactively labelled oligonucleotides specific to that region. Nascent DNA fractions with a wider range of sizes gave a positive signal when

probed with an oligonucleotide deriving from a site in close proximity to the origin. In contrast, probes which mapped further away from the origin could only hybridise to DNA amplified from template representing fractions that were longer in size.

These findings offer a useful insight into the organisation of replication in the MHC. The number of origins that is expected to exist within the MHC based on current models, about 40 for ~3.8 Mb, is likely to be surpassed by the actual number. The existence of background signal in the results of the array experiments presents an obstacle in obtaining precise information on the number of initiation sites. Application of data smoothing approaches such as the "Running mean" reduces the extreme values in the sample population as well as the number of features with background signal. An estimation based on the number of peaks present after smoothing (Figure 5.7) indicates that the number of origins in the region could be 60 or more. This inference makes the assumption that peaks on the figure represent sites of initiation, provided that those sites would be most enriched in nascent DNA. The class III region has a large number of peaks suggesting that it could be more densely-populated by initiation sites than other regions of the genome. The extreme gene-richness of the class III region could be significant, in view of previous observations that certain origins of replication map to promoter regions of genes (Vassilev and Johnson 1990; Giacca, Zentilin et al. 1994; Taira, Iguchi-Ariga et al. 1994; Keller, Ladenburger et al. 2002; Ladenburger, Keller et al. 2002; Paixao, Colaluca et al. 2004).

A comparison between the replication timing profile generated by the experiments described in Chapter 4 (Figure 4.4), and the data on replication initiation in the MHC presented here, suggests that there is overall agreement in the findings of the two methods (Figure 5.9). Although the two plots have a similar trend, the replication timing profile has fewer peaks than that of nascent DNA enrichment. This observation suggests that the method used to examine replication timing is not as sensitive in origin identification as the strategy implemented in this chapter. Notably, the trend-

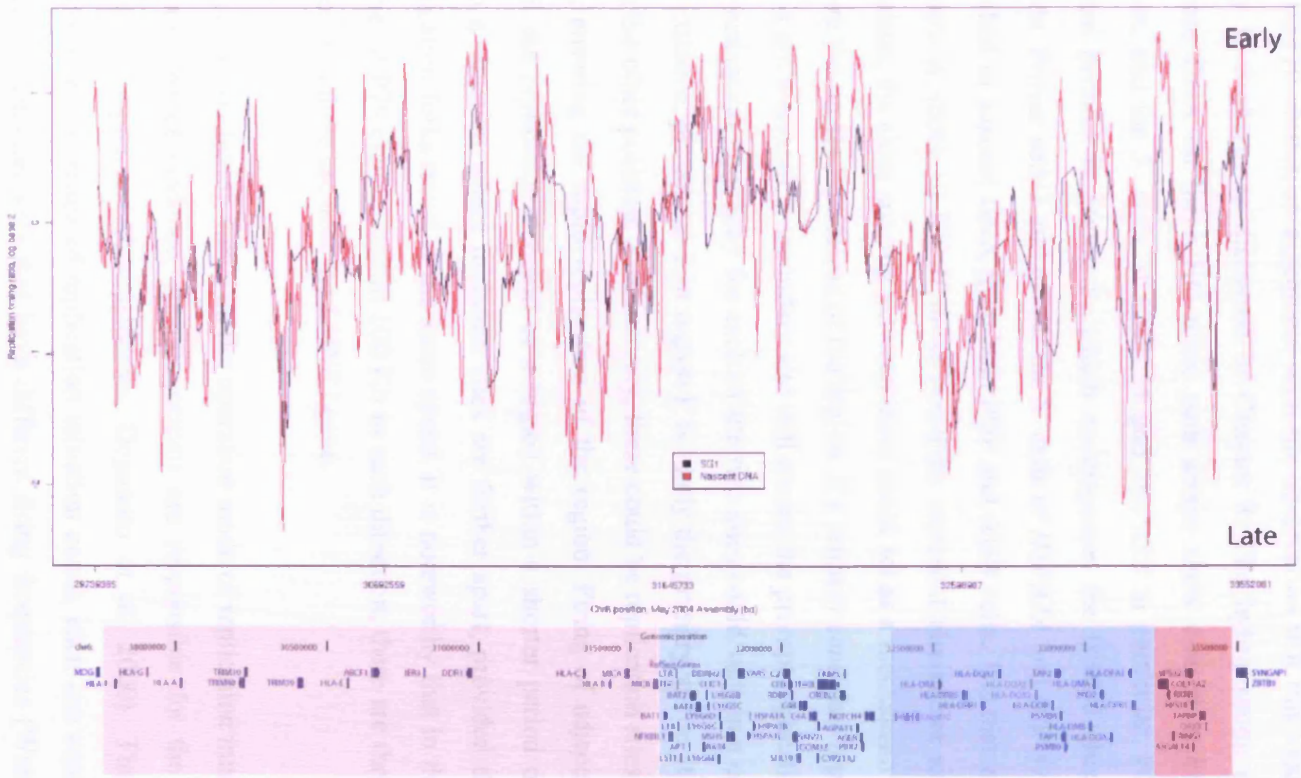


Figure 5.9 Replication timing vs Nascent DNA enrichment.

The "Running mean"-smoothed profiles for replication (black line) and nascent DNA enrichment (red line) are shown together. Positions nearer the top of the graph indicate earlier replication or higher levels of nascent DNA. The two lines have a similar trend, though the red line has more extreme values than the black line. It also has more discrete peaks. Background highlighting as per page 49.

line of nascent DNA enrichment contains more extreme values than the replication timing profile, which could be a consequence of the isolation procedure.

The findings of this study suggest that there is at least one origin of replication within the HSP70 cluster, in agreement with the observation that this locus replicates very early in S-phase, as discussed in Chapter 3. The hybridisation experiments with nascent DNA on the MHC tiling path arrays show consistent enrichment at this region, and the 5′ ends of *HSPA1A* and *HSPA1B* in particular. PCR analysis with several primers in region F, which encompasses the HSP70 cluster, gave similar results. Primer sets 3 and 6/7, at the 5′ ends of *HSPA1A* and *HSPA1B*, were the most enriched in nascent DNA from both PGF and AHB cells. The distance between these primers is about 10 Kb. If those positions represent alternative sites of replication initiation, the close spacing between them could act as a mechanism of redundancy to ensure the timely replication of the region. If a primary candidate replication initiation site is not activated, a secondary site will ensure the prompt replication of that region. The activation frequency for each of the two sites could vary in a cell population. In this instance, primer set 3 in region F is likely the primary site as it is more enriched than the other position. Alternatively, these could be replication sites that fire in every cell, ensuring the rapid replication of the region. Firing of adjacent origins would fulfil the replication demand of a region within a shorter period of time compared with a situation where initiation sites are further apart, provided that in both cases replication forks travel at the same speed. It is noteworthy that in the neighbourhood of the HSP70 cluster, within 100 Kb in each direction, there are three times as many genes than there are near the *LMNB2* gene.

There are conflicting views on the operation mode of replication initiation. According to one line of evidence, specific origins are responsible for the replication of a particular replicon (Abdurashidova, Deganuto et al. 2000). The opposing view supports the existence of replication initiation zones, each one consisting of several potential initiation sites that have different firing frequencies (Wang, Dijkwel et al. 1998; Dijkwel, Wang et al. 2002). The findings from this study suggest that both models can co-exist. The single origin model applies in some instances, as in the case of region M where primer set 3 is much more enriched than the remaining primer sets.

The zone model, on the other hand, could apply in the case of region F, as discussed above.

Chapter 6

Final Discussion

6.1 Summary

In this thesis I have investigated the timing of DNA replication in the human MHC, using different approaches, and also searched for sites of replication initiation within this region.

The replication timing was first analysed at ten loci from the MHC class II and III regions (Chapter 3). Immunoprecipitation of BrdU-enriched DNA from cells sorted into different cell cycle sub-populations followed by southern hybridisations on amplified locus-specific products was used to examine replication in three different cell types. In the B-lymphoblastoid cell line AHB, all ten loci were found to replicate at the earliest measured times, as BrdU enriched DNA was most abundant in the S1 fraction. Using the same approach, the ten MHC loci were found to replicate slightly later in MRC5 fibroblasts, with BrdU-enriched DNA being most abundant in the S1 and S2 sub-populations. Lastly, the myeloid leukaemia cell line HL60 was found to have a similar replication programme to the fibroblasts. To map replication timing at the ten MHC loci more precisely, further analysis was performed using more uniform populations of HL60 cells. The cells were synchronised at the G1/S-phase boundary by aphidicolin and samples collected at different timepoints after release from block. The samples were then used to flow-sort homogeneous cell populations, followed by southern hybridisation on amplified locus-specific products. This allowed replication timing at the different loci to be distinguished more clearly. The ten MHC loci were found to replicate at different timepoints using this technique. The MHC class III loci replicated very early, within the first 3 hours of S-phase, while classical class II loci replicated later. The earliest replicating MHC locus was HSPA1 and the latest replicating was HLA-DRA. Taken together, these findings suggest that the ten MHC loci replicate within the first half of S-phase, but that the class II and class III regions have different features.

Replication timing in the MHC was also examined using an array approach (Chapter 4). Following flow-sorting of cell cycle fractions, DNA from AHB cells in S-phase was co-hybridised with G1 DNA on high resolution tiling path arrays covering ~3.8 Mb of the MHC. The classical MHC class II region was found to replicate later than the adjacent regions. These data were in agreement with the findings described in Chapter 3 as well as those of published studies. Statistical analyses showed a weak positive correlation between replication timing and various features of the genome such as GC content, exon content and CpG island density.

Finally, I searched for sites of replication initiation in the MHC region (Chapter 5). Nascent DNA was isolated on the basis of size fractionation of single stranded DNA extracted from nuclei. Nascent DNA samples isolated from B lymphoblastoid AHB and PGF cells were then co-hybridised with G1 DNA onto the MHC tiling path array. Overall, the two cell lines gave similar results. The synoptic array data indicated that there could be as many as 60 origins in the MHC region. A weak positive correlation was found between GC content and nascent DNA abundance. Also, the very gene-rich MHC class III appears to have a large number of replication origins as it had the highest number of peaks. Further analysis at 13 MHC loci by PCR also showed similarities in nascent DNA enrichment between the two cell lines. Notably, however, the two sets of findings should be interpreted with regard to the methodological approach taken, as the array and PCR data were not always in agreement with each other. Taken together the data presented suggest that replication initiation in the MHC could take place either via origins of replication or zones of replication initiation.

6.2 Wider implications of the findings

The findings described in the three previous chapters provide detailed information about DNA replication in the human MHC region. The significance of transcription and genomic features, have been discussed with respect to timing of replication and sites of replication initiation. Our current understanding of MHC genetics in conjunction with observations from various replication studies provide some exciting ideas to consider.

The MHC has been extensively studied for linkage disequilibrium (LD) – the non-random association of alleles into haplotypes (also discussed in section 4.3) – as the highly polymorphic classical MHC class I, class II, and class I-related MIC genes are inherited as linked sets of alleles (reviewed by Carrington 1999). Given that the human MHC contains several gene clusters, this type of organisation could serve as a means of preserving combinations of genes that work efficiently and could also facilitate their transcriptional regulation (reviewed by: Beck and Trowsdale 2000; Trowsdale 2002; Kelley, Walter et al. 2005).

Recently, detailed sequence analysis for two individuals identified high levels of conservation within a specific portion of the MHC class II region, amongst a background of extreme variation (Traherne, Horton et al. 2006). Those findings suggested that the conserved region had been subjected to genetic exchange in the early stages of human evolution. The spread in the population of particular haplotypes, which are associated with better-adapted immune function, could also be explained by those results.

Population processes like selection, genetic drift, population bottlenecks, migration and admixture, or genetic events such as recurrent mutation and frequency of recombination are among the factors that can influence the level of LD (Jeffreys, Holloway et al. 2004; Kauppi, Jeffreys et al. 2004). To understand haplotype evolution better, Jeffreys and colleagues have studied the effects of meiotic recombination in the MHC by examining crossover hotspots and their frequencies in the MHC class II region. They showed that in a 292 Kb segment of the class II region, LD "block" structure is associated with the presence of meiotic recombination hotspots. Three regions show crossover clustering: the *RING3–DOA* interval, the *TAP2* region, and the *DQB1–DQB3* interval (Jeffreys, Kauppi et al. 2001; Kauppi, Stumpf et al. 2005). Recombination is thought to be necessary for correct chromosome segregation at the first meiotic division. However, this process is held back in order to minimise the dispersal of favourable linkage groups, in line with the suggestion made by Traherne *et al.* and others, and to foster genome stability. The non-random distribution of crossovers provides a means of achieving this balance (Kauppi, Jeffreys et al. 2004).

Although gene order in the human and mouse MHC is conserved, crossover hotspot locations apparently are not. This suggests that hotspot location is not dictated by the constituent genes within a haplotype or that the two species have different requirements for particular genes in a haplotype. It also suggests that the recombination hotspots in the region have a shorter life span than the time since human-mouse divergence (Kauppi, Jeffreys et al. 2004).

The *HBB* gene region of replication initiation has been found to overlap with a putative recombination hotspot, based on evidence from a study of crossover breakpoints in three families (Smith, Ho et al. 1998). Similar findings have been reported in a more recent study (Wall, Frisse et al. 2003). A link between meiotic recombination and replication initiation had initially been established in *Saccharomyces cerevisiae* (Borde, Goldman et al. 2000). The reported affinity of the origin-ubiquitous ORC complex for AT-rich DNA (Vashee, Cvetic et al. 2003), which can be easily unwound, may create a chromatin structure prone to recombination.

Our findings show that the MHC replicates in the first half of S-phase, and that it contains earlier and later replicating parts. An investigation into the potential relationship between the replication programme and the LD "block" structure or crossover hotspot positioning in the MHC is of great interest. Selection pressure to retain certain combinations of alleles is likely to influence DNA metabolism in that region and possibly effect replication organisation.

Watanabe *et al.* examined replication timing at 11q and 21q, using a PCR-based approach similar to that in Chapter 3 of this thesis. Early to late replication transitions occurred primarily at or proximal to GC% transitions, while SNP frequency was high in the late-replicating and replication-transition regions. Intriguingly, cancer and other disease-related genes resided in timing-switch regions (Watanabe, Fujiyama et al. 2002).

Four different studies, have used FISH to investigate replication timing at various loci in samples from unaffected individuals and patients suffering from i) chronic myeloid leukaemia/lymphoma (Amiel, Litmanovitch et al. 1998); ii) renal cell carcinoma

(Dotan, Dotan et al. 2000); iii) chronic lymphocytic leukaemia (Amiel, Elis et al. 2001); and iv) neurofibromatosis type 1 (Reish, Orlovski et al. 2003). In all these studies, homologous loci that replicated with synchrony in normal cells were found to replicate asynchronously at a higher rate in cells from affected individuals.

Other studies have also identified alterations in the replication timing programme that accompany activation of the fragile site FRA3B (Wang, Darling et al. 1999), translocation of *BCL2* from chromosome 18 to 14 (Sun, Wyatt et al. 2001), and deletion of 22q11 in the DiGeorge and Velocardiofacial syndromes (D'Antoni, Mattina et al. 2004).

These findings indicate the existence of an association between disease and the perturbation of the normal replication programme. Dotan *et al.* have suggested that the alteration in replication as detected by FISH applied to lymphocytes, offers a potential test for cancer identification (Dotan, Dotan et al. 2000). Could this approach be applied to monitor disease association for other genomic loci and diseases? Watanabe *et al.* propose that genome-wide assessment of replication timing serves as an efficient strategy for identifying disease-related genes since increased DNA damage is reported to occur in replication-transition regions (Watanabe, Fujiyama et al. 2002).

The implications of the findings from these studies should be considered with reference to the MHC, in order to examine whether genes associated with disease reside in regions with asynchronous replication or in replication timing transitions.

6.3 Future directions

The findings presented in this thesis have provided insights into various aspects of replication organisation in the MHC and have generated new questions.

Firstly, it is important to recognise ways by which these data could be analysed further and perhaps combined with information from other sources to expand our understanding of the organisation of replication in the MHC. Secondly, it is

appropriate to consider approaches through which the experiments presented in this thesis could be improved to strengthen the findings.

Two features that could potentially be examined using the array data presented in this thesis are replication termination sites in the MHC and fork migration rates. Based on the graphical representation of S:G1 ratio (Figure 4.4), where peaks likely correspond to replication initiation sites, the troughs in the same figure should indicate sites of replication termination. As the replication forks from adjacent replicons travel away from the site of initiation, they will eventually merge to terminate replication in that particular genomic locus. Taking this assumption a step further, the time-span between initiation and termination of replication – represented by the slopes from each trough to each peak in Figure 4.4 – could be used to calculate the speed at which DNA synthesis takes place, the fork migration rate. One of the two parameters relevant to this calculation, the genomic distance covered by each fork could be extracted from these data. The second parameter, the time required by the replication fork to perform this task, can be inferred in reference to other studies that have established absolute relationships for replication timing across the entire genome (Woodfine, Fiegler et al. 2004). In addition, subsequent analysis might establish the structure of replicons in the MHC, as previous studies have described asymmetric replicons in other genomic regions (Verbovaia and Razin 1997).

The association of the genome with the proposed nuclear structure, generally called the nuclear matrix, scaffold or skeleton, is mediated by MARs or SARs. A bipartite sequence element, the MAR/SAR recognition signature (MRS), initially described by van Drunen and colleagues (van Drunen, Sewalt et al. 1999) has been used in bioinformatics analyses to identify potential MARs. Previously, the collaboration between the Human Cytogenetics Laboratory and Dr Stephan Beck's team at the Welcome Trust Sanger Institute, identified five MARs in the *LMP/TAP* gene cluster in the human MHC (Donev, Horton et al. 2003). The laboratory is currently conducting experiments to identify MARs across the entire MHC with the high resolution array used in this thesis. Several reports have suggested that the nuclear matrix plays a vital role for DNA replication (reviewed by Anachkova, Djeliova et al. 2005). It is thus important to examine MAR positions with respect to replication timing in the MHC. Also of interest is the proposed relationship between MARs and

sites of replication initiation. There is evidence that active origins of replication are associated with the nuclear matrix throughout the cell cycle (Carri, Micheli et al. 1986). Therefore, another issue is whether the proposed association of chromatin with the nuclear matrix affects the speed of the replication fork.

In addition, our laboratory has investigated gene expression in the AHB and PGF B-lymphoblastoid cells, as well as the MRC5 fibroblasts that have been used in this study (Dr Tim Forshew, personal communication). An examination of the replication data from this thesis together with the expression data that are now available may provide further insights into the organisation of replication in the MHC. Statistical analyses to test the relationship between replication and transcription should also be performed.

It is important to consider various issues that arise from the methodology employed in this thesis as well as the approaches used to analyse the data. Most of the following relate to the array experiments, as this is a relatively recent technique.

First and foremost, the MHC arrays to be used in future experiments for replication timing and initiation of replication should have suitable controls. For replication timing the inclusion of loci that are known to replicate very early and very late in S-phase could be used to get an absolute timing of replication for the MHC. It is also necessary to include known origins of replication as these would boost confidence in the investigation of novel initiation sites.

Another point related to the methodological approach of the array experiments involves the use of Cot1 DNA. Pre-hybridisation with human Cot1 DNA is performed to block repetitive sequences prior to application of the Cy3/Cy5-labelled test and control DNA samples. If this step is omitted highly-abundant SINE and LINE elements could bind non-specifically and thus provide misleading results. However, Newkirk and colleagues discuss that this practice introduces bias in the results (Newkirk, Knoll et al. 2005). This ambiguity raises questions about the potential of methodological artefacts in the findings. Blocking-out repetitive sequences from the analysis would restrict the ability to identify potential sites of replication initiation, and create a bias towards non-repetitive gene-related loci, such as exons. A revised

approach should therefore be considered to eliminate such bias. Application of some of the approaches outlined in section 4.1 would re-enforce the validity of the results described in this thesis.

Finally, it is important to consider new directions and approaches that will further our knowledge. To understand the association between the temporal and mechanistic aspect of replication organisation in the MHC, a comprehensive investigation into the binding sites of proteins of the replication machinery must be carried out. Additionally, further array hybridisation experiments need to be performed with MRC5 fibroblasts, before and after induction by IFN-γ, as well as with myeloid leukaemia HL60 cells and PGF lymphoblastoid cells in order to consolidate our knowledge of replication timing and sites of replication initiation for these cells. All these findings will supply a substantial body of information to help us understand how replication timing and origin usage differ in various cell types.

The experiments described above in association with further experiments that will investigate the organisation of chromatin in the MHC, including histone post-translational modifications and other chromatin attributes, will clarify the role that chromatin structure plays in replication timing and in origin selection.

6.4 Concluding remarks

It is, of course, of interest to grasp the defining factors of replication timing and origin usage, such as the GC content, gene density, gene expression, chromatin status, and other features. However, it is far more compelling to unravel how this element of DNA metabolism is encapsulated in the magnitude of a genome and "safely" carried over by selection through the evolutionary process.

Appendix A

Further information on proteins that affect chromatin structure (Adkins, Watts et al. 2004) and references therein).

As previously mentioned, LH are linker DNA binding proteins. They bind at the nucleosome DNA entry/exit point thus reducing the mobility and sliding of the histone octamer. Although the exact structural role of histone H1 and its variants within the chromatin fibre remains controversial, interaction of H1 with nucleosomes increases the nuclease protected area from 146 to 165 base pairs. The C-terminal tail of H5, another LH, is believed to be capable of acquiring a conformation that facilitates chromatin condensation by neutralising charges on the DNA backbone.
Overall, LH stabilise the higher-order chromatin structure and act as general repressors by decreasing DNA accessibility to transcription factors.

HMGB is a group of proteins containing the HMG box which binds DNA in a non-sequence specific manner. They seem to recognise specific DNA structures, and bind preferentially to the minor groove of locally distorted B-form DNA. HMGB proteins share binding locations with H1; this implies a putative competition for nucleosomal association. Despite their classification as DNA binding proteins, HMGB-1 and -2 have been shown to interact with various protein partners. HMGB-2 facilitates binding of a transcription factor, Oct-1, to its target sequence. It was recently demonstrated to physically interact with nucleosome assembly protein SET which is potentially involved in the transcription activation and DNA repair.

HMGN-1 and -2 bind nucleosomes as homodimers. Their interaction with histones H2B and H3 has been associated with chromatin decondensation properties. RNA polymerase II transcriptional activation appears to require the presence of HMGN-1 or -2. Depletion of H1 is accompanied by a simultaneous enrichment in HMGN, as these two groups compete for chromatin binding sites. Furthermore, H1 residence time on chromatin was shown to decrease when the local concentration of HGMN proteins increased, promoting the unfolding of chromatin fibres. The transition from repressed to transcriptionally active chromatin is thought to be the result of a dynamic equilibrium between multiple HMGs.

MENT, the Myeloid and Erythroid Nuclear Termination stage-specific protein is a member of the Serpin protein family of serine protease inhibitors. Apart from DNA, MENT binds nucleosomes independently of histone-tails. It induces chromatin compaction by forming a "zipper-like" structure thus bridging DNA molecules.
This is achieved by bringing linker DNA entry/exit regions in close proximity and then dimerising with a neighbouring MENT molecule.

HP1 is a highly conserved protein associated with heterochromatin. It was initially described in *Drosophila melanogaster* and has three mammalian isoforms, HP1α, -β and -γ. HP1α and β are predominantly localised to centromeric heterochromatin while HP1γ is localised in both heterochromatin and euchromatin. Each of the three isoforms has been found to interact with different partners like Suv39H1 - a histone H3 methyltransferase, BRG1 - a chromatin remodeller, methyltransferases DNMT1

and DNMT3, the retinoblastoma protein and human TAFII130. These interactions suggest a role for HP1 in both gene activation and silencing.

MeCP2 is a member of the methyl CpG-binding domain (MBD) family. It associates with chromatin in densely methylated CpG-rich pericentromeric regions in mouse cells, as well as in major satellite DNA. It also acts as a global transcription repressor. MeCP2-mediated repression is reportedly alleviated by TrichoStatin A, a histone deacetylase inhibitor. Histone De-Acetylase (HDAC) activity was observed when the transcription co-repressor Sin 3A was co-purified with MeCP2. Interestingly, MeCP2 binding to DNA and the ability to condense chromatin do not require DNA methylation.

Appendix B

Further information on the biology of complement system (Sim and Tsiftsoglou 2004) and references therein)

The classical pathway can be activated through binding of the $C1qr_2s_2$ complex to the Fc region of immunoglobulin (Ig) M and certain subclasses of IgG molecules. Consequently, C4 and C2, which are substrates of the C1 complex, are cleaved at specific residues to produce C4a, C4b, C2a and C2b. Their catalytic potential is activated when they form C4b2a, a molecule often referred to as C3 convertase, due to its ability to convert the C3 proenzyme into its active form. The smaller component, C4a, as well as other proteolytic cleavage products of the complement pathway, is an anaphylatoxin. Anaphylatoxins are mediators of inflammation. Some of their main functions include binding to granulated leukocytes to induce release of histamine and other pharmacological substances, induction of smooth-muscle contraction and increased vascular permeability.

The lectin pathway is activated by binding of mannan-binding lectin (MBL) to microbial cell walls. A C1-like molecule is involved in the formation of the C3 convertase, C4b2a, as in the classical pathway.

The alternative pathway becomes activated by foreign cell-surface components. C3 can spontaneously hydrolyse to yield C3a, an anaphylatoxin, and C3b that binds to microbial surfaces and retains its activity. It forms a complex with Factor B, which is cleaved in a reaction catalysed by Factor D, releasing the Ba fragment while the Bb fragment remains bound to C3b. The resulting C3bBb complex has C3 convertase activity, and is stabilised by the serum protein properdin. Interaction with one further C3b subunit results to the formation of C3bBb3b which is a C5 convertase. The same principle applies to C4b2a which assumes C5 convertase activity upon binding of C3b, resulting in a C4b2a3b complex.

Appendix C

The information below has been extracted from the first page of the ".gal" file.

```
ATF             1
      30        5
Type=GenePix ArrayList V1.0
BlockCount=24
BlockType=0
Block1=2000, 19020, 200, 19, 220, 20, 220
Block2=6500, 19020, 200, 19, 220, 20, 220
Block3=11000, 19020, 200, 19, 220, 20, 220
Block4=15500, 19020, 200, 19, 220, 20, 220
Block5=2000, 23520, 200, 19, 220, 20, 220
Block6=6500, 23520, 200, 19, 220, 20, 220
Block7=11000, 23520, 200, 19, 220, 20, 220
Block8=15500, 23520, 200, 19, 220, 20, 220
Block9=2000, 28020, 200, 19, 220, 20, 220
Block10=6500, 28020, 200, 19, 220, 20, 220
Block11=11000, 28020, 200, 19, 220, 20, 220
Block12=15500, 28020, 200, 19, 220, 20, 220
Block13=2000, 32520, 200, 19, 220, 20, 220
Block14=6500, 32520, 200, 19, 220, 20, 220
Block15=11000, 32520, 200, 19, 220, 20, 220
Block16=15500, 32520, 200, 19, 220, 20, 220
Block17=2000, 37020, 200, 19, 220, 20, 220
Block18=6500, 37020, 200, 19, 220, 20, 220
Block19=11000, 37020, 200, 19, 220, 20, 220
Block20=15500, 37020, 200, 19, 220, 20, 220
Block21=2000, 41520, 200, 19, 220, 20, 220
Block22=6500, 41520, 200, 19, 220, 20, 220
Block23=11000, 41520, 200, 19, 220, 20, 220
Block24=15500, 41520, 200, 19, 220, 20, 220
Supplier=BioRobotics
ArrayerSoftwareName=TAS Application Suite (MicroGrid II)
ArrayerSoftwareVersion=2.6.0.3
```

Block	Column	Row	ID	Name
1	1	1	CY3_102	unknown
1	1	2	stSG1159313_REV	unknown
1	1	3	mtp_S6C_2_558c22_REV	mtp_S6C_2_558c22_REV
1	1	4	mtp_S6C_2_65d23_REV	mtp_S6C_2_65d23_REV
1	1	5	mtp_6S17_2_605k18_REV	mtp_6S17_2_605k18_REV
1	1	6	mtp_S6C_3_141l24_FOR	mtp_S6C_3_141l24_FOR
1	1	7	mtp_S6C_2_313h24_FOR	mtp_S6C_2_313h24_FOR
1	1	8	mtp_S6A_2_593o01_FOR	mtp_S6A_2_593o01_FOR
1	1	9	mtp_S6A_2_80h01_FOR	mtp_S6A_2_80h01_FOR
1	1	10	mtp_6S17_2_174j09_FOR	mtp_6S17_2_174j09_FOR
1	1	11	CY3_102	unknown
1	1	12	stSG1159313_REV	unknown
1	1	13	mtp_S6C_2_558c22_REV	mtp_S6C_2_558c22_REV
1	1	14	mtp_S6C_2_65d23_REV	mtp_S6C_2_65d23_REV
1	1	15	mtp_6S17_2_605k18_REV	mtp_6S17_2_605k18_REV
1	1	16	mtp_S6C_3_141l24_FOR	mtp_S6C_3_141l24_FOR

Appendix D

The information below has been extracted from the first page of the "ids.txt" file.

```
tile_path: 107420 1865 -2664 840 29739385 29741249 6S17_2-317p19.q1kw
tile_path: 107433 2660 -3062 1779 29741918 29744577 S6A_2-77l14.p1kk
tile_path: 107444 2575 -2574 1500 29744676 29747250 6S17_2-612l05.q1ka
tile_path: 107450 2997 -2996 1570 29746874 29749870 6S17_2-476d07.q1kw
tile_path: 107462 2528 -2527 1775 29749779 29752306 S6A_2-58b24.q1kk
tile_path: 107467 2794 -2793 1551 29751462 29754255 6S17_2-791d16.p1kw
tile_path: 107474 2467 -2466 1806 29753458 29755924 S6C_2-527b07.p1ka
tile_path: 107480 915 -2409 834 29755213 29757127 6S17_2-247p20.p1kw
tile_path: 107487 2797 -2796 1148 29757489 29760285 S6C_2-310l18.p1kw
tile_path: 107497 2067 -2328 2106 29760268 29762334 S6A_2-161l04.q1k
tile_path: 107504 2760 -2759 1589 29761508 29764267 S6A_2-44e16.q1k
tile_path: 107505 2875 -2874 1604 29764237 29767111 6S17_2-225m22.p1kw
tile_path: 107509 2395 -2394 1445 29766552 29768946 S6A_2-550n14.q1kw
tile_path: 107513 2376 -2375 1503 29767842 29770217 S6C_2-295c12.p1kw
tile_path: 107517 2335 -2610 1363 29770200 29772534 S6A_2-271c01.p1k
tile_path: 107526 2635 -2634 1402 29772399 29775033 S6A_2-31n18.p1k
tile_path: 107536 2663 -2662 1522 29774950 29777612 S6C_2-342g03.p1kw
tile_path: 107544 2400 -2589 974 29777595 29779994 S6C_2-600c01.q1kw
tile_path: 107553 2532 -2563 1365 29779522 29782053 6S17_2-677d21.q1kw
tile_path: 107563 2403 -2402 1544 29781700 29784102 6S17_2-432b11.q1kw
tile_path: 107572 2540 -2547 1669 29784008 29786547 S6C_2-33f20.q1kkw
tile_path: 107580 2471 -2656 1419 29786066 29788536 S6A_2-388e21.q1kw
tile_path: 107591 2431 -2430 1408 29788550 29790980 S6A_2-84i07.p1kk
tile_path: 107598 1989 -1988 803 29790708 29792696 6S17_2-561c08.p1kw
tile_path: 107606 5832 -5831 1724 29792469 29798300 S6C_3-101a09.p1kw
tile_path: 107611 2811 -2810 1708 29797138 29799948 S6C_2-557i06.p1kw

gap: 29802662 29800446 S6A_2-697f03.q1kw S6A_2-697f03.q1kw

tile_path: 107612 2217 -2216 1411 29800446 29802662 S6A_2-697f03.q1kw
tile_path: 107613 1824 -1823 783 29802604 29804427 S6A_2-592e18.q1kw
tile_path: 107616 1092 -2091 948 29804125 29806216 6S17_2-714f01.q1kw
tile_path: 107621 2723 -2722 1668 29805657 29808379 S6C_2-296i15.p1kw
tile_path: 107625 2174 -2173 1618 29807664 29809837 S6A_2-424g15.q1kw
tile_path: 107629 2480 -2479 1287 29808891 29811370 S6C_2-497m09.q1kw
tile_path: 107638 2368 -2391 1627 29811013 29813380 S6A_2-372b12.q1kw
tile_path: 107641 2526 -2525 1411 29812577 29815102 6S17_2-614l12.p1ka
tile_path: 107646 2338 -2452 1457 29814777 29817114 S6C_2-242e07.p1kw
tile_path: 107650 2681 -2680 894 29817132 29819812 S6C_2-704i08.p1kw
tile_path: 107660 2716 -2715 2051 29819808 29822523 6S17_2-471h23.q1kw
tile_path: 107665 1962 -2282 1449 29821557 29823518 S6C_2-722b13.p1kw

gap: 29826356 29823989 S6A_2-373m17.p1kw S6A_2-373m17.p1kw

tile_path: 107666 2368 -2367 1281 29823989 29826356 S6A_2-373m17.p1kw
```

Appendix E

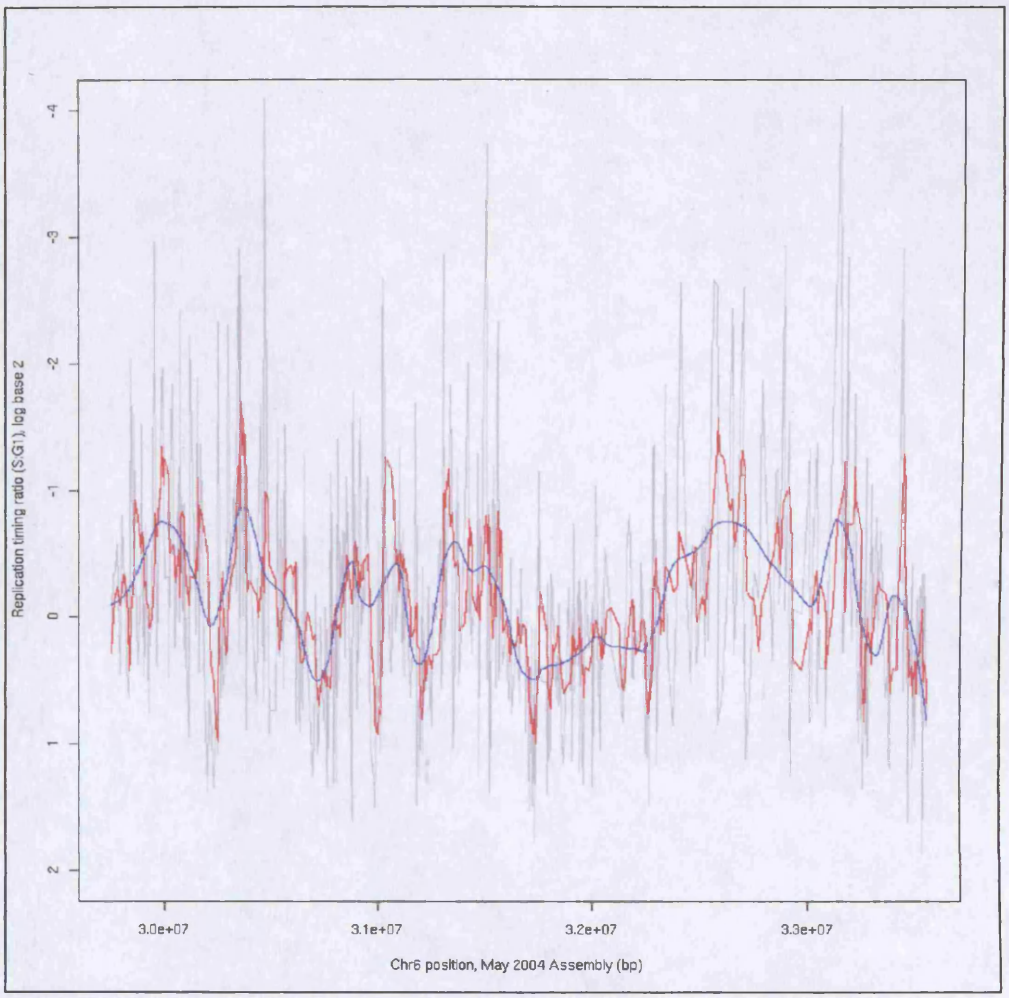
Further information on the studies that examined replication timing at the loci used as replication timing controls.

The glycogen phosphorylase gene (*PYGM*) covers ~15 kb on the q arm of chromosome 11. It encodes an enzyme that catalyses and regulates the breakdown of glycogen to glucose-1-phosphate. *PYGM* is expressed in several tissues including cells of the immune, nervous and muscular system (Shmueli, Horn-Saban et al. 2003). The human β-globin gene (*HBB*) spans ~1.6 kb on the p arm of chromosome 11. The encoded polypeptide is a component of adult haemoglobin and is expressed primarily in cells of erythroid lineage (Epner, Forrester et al. 1988; Dhar, Skoultchi et al. 1989).

Using a FISH approach in an asynchronous cell population, Selig and co-workers investigated the replication timing of these two genes simultaneously in different cell lines (Selig, Okumura et al. 1992). They found that HL60 and Manca, a myeloid and a lymphoma cell line respectively, have a higher proportion of cells that replicate the *PYGM* before the *HBB* gene locus. Conversely, the same observation was not made with the erythroleukaemia cell line K562, for which both loci were found to replicate at similar times early in S-phase. Another study used a similar approach to examine replication timing for the two loci in three lymphoblastoid cell lines, in which *PYGM* was found to replicate earlier than *HBB* (Smith and Higgs 1999). It also showed that the alpha globin gene (*HBA*), at 16p, replicates early in different cell lines, regardless of expression.

Lamin B2 (*LB2*), replicates immediately after the onset of S-phase according to a nascent DNA analysis (Giacca, Zentilin et al. 1994).

Appendix F



Application of smoothing approaches to Replication timing, $\log_2$ ratio values (Inverted image of Figure 4.4).

Regions with early replication face towards the bottom of the page and those with later replication towards the top.

References

Abdurashidova, G., M. Deganuto, et al. (2000). "Start sites of bidirectional DNA synthesis at the human lamin B2 origin." Science **287**(5460): 2023-6.

Adkins, N. L., M. Watts, et al. (2004). "To the 30-nm chromatin fiber and beyond." Biochim Biophys Acta **1677**(1-3): 12-23.

Aladjem, M. I., M. Groudine, et al. (1995). "Participation of the human beta-globin locus control region in initiation of DNA replication." Science **270**(5237): 815-9.

Aladjem, M. I. and G. M. Wahl (1997). "Mapping replication fork direction by leading strand analysis." Methods **13**(3): 281-92.

Altman, A. L. and E. Fanning (2001). "The Chinese hamster dihydrofolate reductase replication origin beta is active at multiple ectopic chromosomal locations and requires specific DNA sequence elements for activity." Mol Cell Biol **21**(4): 1098-110.

Amiel, A., A. Elis, et al. (2001). "The influence of cytogenetic aberrations on gene replication in chronic lymphocytic leukemia patients." Cancer Genet Cytogenet **125**(2): 81-6.

Amiel, A., T. Litmanovitch, et al. (1998). "Temporal differences in replication timing of homologous loci in malignant cells derived from CML and lymphoma patients." Genes Chromosomes Cancer **22**(3): 225-31.

Anachkova, B., V. Djeliova, et al. (2005). "Nuclear matrix support of DNA replication." J Cell Biochem **96**(5): 951-61.

Andersson, G., A. C. Svensson, et al. (1998). "Retroelements in the human MHC class II region." Trends Genet **14**(3): 109-14.

Androlewicz, M. J., K. S. Anderson, et al. (1993). "Evidence that transporters associated with antigen processing translocate a major histocompatibility complex class I-binding peptide into the endoplasmic reticulum in an ATP-dependent manner." Proc Natl Acad Sci U S A **90**(19): 9130-4.

Anglana, M., F. Apiou, et al. (2003). "Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing." Cell **114**(3): 385-94.

Antequera, F. and A. Bird (1999). "CpG islands as genomic footprints of promoters that are associated with replication origins." Curr Biol **9**(17): R661-7.

Azuara, V., K. E. Brown, et al. (2003). "Heritable gene silencing in lymphocytes delays chromatid resolution without affecting the timing of DNA replication." Nat Cell Biol **5**(7): 668-74.

Bakke, O. and B. Dobberstein (1990). "MHC class II-associated invariant chain contains a sorting signal for endosomal compartments." Cell **63**(4): 707-16.

Bartek, J., C. Lukas, et al. (2004). "Checking on DNA damage in S phase." Nat Rev Mol Cell Biol **5**(10): 792-804.

Batchelor, J. R. and V. C. Joysey (1969). "Influence of HL-A incompatibility on cadaveric renal transplantation." Lancet **1**(7599): 790-2.

Beck, S. and J. Trowsdale (2000). "The human major histocompatability complex: lessons from the DNA sequence." Annu Rev Genomics Hum Genet **1**: 117-37.

Becker, P. B. and W. Horz (2002). "ATP-dependent nucleosome remodeling." Annu Rev Biochem **71**: 247-73.

Bell, S. P. (1995). "Eukaryotic replicators and associated protein complexes." Curr Opin Genet Dev **5**(2): 162-7.

Bernardi, G. (1995). "The human genome: organization and evolutionary history." Annu Rev Genet **29**: 445-76.
Bernardi, G. (2000). "Isochores and the evolutionary genomics of vertebrates." Gene **241**(1): 3-17.
Bernardi, G., B. Olofsson, et al. (1985). "The mosaic genome of warm-blooded vertebrates." Science **228**(4702): 953-8.
Biamonti, G., M. Giacca, et al. (1992). "The gene for a novel human lamin maps at a highly transcribed locus of chromosome 19 which replicates at the onset of S-phase." Mol Cell Biol **12**(8): 3499-506.
Bickmore, W. A. and A. D. Carothers (1995). "Factors affecting the timing and imprinting of replication on a mammalian chromosome." J Cell Sci **108 (Pt 8)**: 2801-9.
Bilyeu, K. and A. C. Chinault (1998). "Replication timing properties across the pseudoautosomal region boundary and cytogenetic band boundaries on human distal Xp." Chromosoma **107**(2): 105-12.
Bird, A. P. (1986). "CpG-rich islands and the function of DNA methylation." Nature **321**(6067): 209-13.
Bjorkman, P. J., M. A. Saper, et al. (1987). "Structure of the human class I histocompatibility antigen, HLA-A2." Nature **329**(6139): 506-12.
Blanchong, C. A., E. K. Chung, et al. (2001). "Genetic, structural and functional diversities of human complement components C4A and C4B and their mouse homologues, Slp and C4." Int Immunopharmacol **1**(3): 365-92.
Blow, J. J., P. J. Gillespie, et al. (2001). "Replication origins in Xenopus egg extract Are 5-15 kilobases apart and are activated in clusters that fire at different times." J Cell Biol **152**(1): 15-25.
Bogenhagen, D. and D. A. Clayton (1977). "Mouse L cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle." Cell **11**(4): 719-27.
Bolzer, A., G. Kreth, et al. (2005). "Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes." PLoS Biol **3**(5): e157.
Borde, V., A. S. Goldman, et al. (2000). "Direct coupling between meiotic DNA replication and recombination initiation." Science **290**(5492): 806-9.
Braud, V., E. Y. Jones, et al. (1997). "The human major histocompatibility complex class Ib molecule HLA-E binds signal sequence-derived peptides with primary anchor residues at positions 2 and 9." Eur J Immunol **27**(5): 1164-9.
Brylawski, B. P., S. M. Cohen, et al. (2004). "Transitions in replication timing in a 340 kb region of human chromosomal R-Band 1p36.1." J Cell Biochem **92**(4): 755-69.
Burhans, W. C., L. T. Vassilev, et al. (1990). "Identification of an origin of bidirectional DNA replication in mammalian chromosomes." Cell **62**(5): 955-65.
Busch, R., C. H. Rinderknecht, et al. (2005). "Achieving stability through editing and chaperoning: regulation of MHC class II peptide binding and expression." Immunol Rev **207**: 242-60.
Canman, C. E. (2001). "Replication checkpoint: preventing mitotic catastrophe." Curr Biol **11**(4): R121-4.
Cardoso, M. C., H. Leonhardt, et al. (1993). "Reversal of terminal differentiation and control of DNA replication: cyclin A and Cdk2 specifically localize at subnuclear sites of DNA replication." Cell **74**(6): 979-92.

Carri, M. T., G. Micheli, et al. (1986). "The relationship between chromosomal origins of replication and the nuclear matrix during the cell cycle." Exp Cell Res **164**(2): 426-36.
Carrington, M. (1999). "Recombination within the human MHC." Immunol Rev **167**: 245-56.
Chess, A., I. Simon, et al. (1994). "Allelic inactivation regulates olfactory receptor gene expression." Cell **78**(5): 823-34.
Cimbora, D. M., D. Schubeler, et al. (2000). "Long-distance control of origin choice and replication timing in the human beta-globin locus are independent of the locus control region." Mol Cell Biol **20**(15): 5581-91.
Cohen, S. M., B. P. Brylawski, et al. (2002). "Mapping of an origin of DNA replication near the transcriptional promoter of the human HPRT gene." J Cell Biochem **85**(2): 346-56.
Cohen, S. M., B. P. Brylawski, et al. (2003). "Same origins of DNA replication function on the active and inactive human X chromosomes." J Cell Biochem **88**(5): 923-31.
Cohen, S. M., E. R. Cobb, et al. (1998). "Identification of chromosomal bands replicating early in the S phase of normal human fibroblasts." Exp Cell Res **245**(2): 321-9.
Collins, F. S., E. S. Lander, et al. (2004). "Finishing the euchromatic sequence of the human genome." Nature **431**(7011): 931-45.
Collins, S. J. (2002). "The role of retinoids and retinoic acid receptors in normal hematopoiesis." Leukemia **16**(10): 1896-905.
Coverley, D., C. Pelizon, et al. (2000). "Chromatin-bound Cdc6 persists in S and G2 phases in human cells, while soluble Cdc6 is destroyed in a cyclin A-cdk2 dependent process." J Cell Sci **113 (Pt 11)**: 1929-38.
Craig, J. M. and W. A. Bickmore (1994). "The distribution of CpG islands in mammalian chromosomes." Nat Genet **7**(3): 376-82.
Cremer, T. and C. Cremer (2006). "Rise, fall and resurrection of chromosome territories: a historical perspective. Part I. The rise of chromosome territories." Eur J Histochem **50**(3): 161-76.
Cremer, T. and C. Cremer (2006). "Rise, fall and resurrection of chromosome territories: a historical perspective. Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present." Eur J Histochem **50**(4): 223-72.
Cremer, T., A. Kurz, et al. (1993). "Role of chromosome territories in the functional compartmentalization of the cell nucleus." Cold Spring Harb Symp Quant Biol **58**: 777-92.
Cresswell, P. (1996). "Invariant chain structure and MHC class II function." Cell **84**(4): 505-7.
Cross, S. H. and A. P. Bird (1995). "CpG islands and genes." Curr Opin Genet Dev **5**(3): 309-14.
Cuny, G., P. Soriano, et al. (1981). "The major components of the mouse and human genomes. 1. Preparation, basic properties and compositional heterogeneity." Eur J Biochem **115**(2): 227-33.
Cuvier, O. and T. Hirano (2003). "A role of topoisomerase II in linking DNA replication to chromosome condensation." J Cell Biol **160**(5): 645-55.

D'Antoni, S., T. Mattina, et al. (2004). "Altered replication timing of the HIRA/Tuple1 locus in the DiGeorge and Velocardiofacial syndromes." Gene **333**: 111-9.

Dai, J., R. Y. Chuang, et al. (2005). "DNA replication origins in the Schizosaccharomyces pombe genome." Proc Natl Acad Sci U S A **102**(2): 337-42.

Dausset, J. (1958). "[Iso-leuko-antibodies.]." Acta Haematol **20**(1-4): 156-66.

Davis, A. F. and D. A. Clayton (1996). "In situ localization of mitochondrial DNA replication in intact mammalian cells." J Cell Biol **135**(4): 883-93.

Dayan, C. M., M. Londei, et al. (1991). "Autoantigen recognition by thyroid-infiltrating T cells in Graves disease." Proc Natl Acad Sci U S A **88**(16): 7415-9.

Deininger, M. H., R. Meyermann, et al. (2002). "The allograft inflammatory factor-1 family of proteins." FEBS Lett **514**(2-3): 115-21.

Delgado, S., M. Gomez, et al. (1998). "Initiation of DNA replication at CpG islands in mammalian chromosomes." Embo J **17**(8): 2426-35.

DePamphilis, M. L. (1993). "Eukaryotic DNA replication: anatomy of an origin." Annu Rev Biochem **62**: 29-63.

DePamphilis, M. L. (1998). "Initiation of DNA replication in eukaryotic chromosomes." J Cell Biochem Suppl **30-31**: 8-17.

DePamphilis, M. L. (1999). "Replication origins in metazoan chromosomes: fact or fiction?" Bioessays **21**(1): 5-16.

DePamphilis, M. L. (2003). "Eukaryotic DNA replication origins: reconciling disparate data." Cell **114**(3): 274-5.

Dermitzakis, E. T., A. Reymond, et al. (2005). "Conserved non-genic sequences - an unexpected feature of mammalian genomes." Nat Rev Genet **6**(2): 151-7.

Dewannieux, M. and T. Heidmann (2005). "Role of poly(A) tail length in Alu retrotransposition." Genomics **86**(3): 378-81.

Dhar, V., D. Mager, et al. (1988). "The coordinate replication of the human beta-globin gene domain reflects its transcriptional activity and nuclease hypersensitivity." Mol Cell Biol **8**(11): 4958-65.

Dhar, V., A. I. Skoultchi, et al. (1989). "Activation and repression of a beta-globin gene in cell hybrids is accompanied by a shift in its temporal replication." Mol Cell Biol **9**(8): 3524-32.

Diehl, M., C. Munz, et al. (1996). "Nonclassical HLA-G molecules are classical peptide presenters." Curr Biol **6**(3): 305-14.

Dijkwel, P. A., S. Wang, et al. (2002). "Initiation sites are distributed at frequent intervals in the Chinese hamster dihydrofolate reductase origin of replication but are used with very different efficiencies." Mol Cell Biol **22**(9): 3053-65.

Donev, R., R. Horton, et al. (2003). "Recruitment of heterogeneous nuclear ribonucleoprotein A1 in vivo to the LMP/TAP region of the major histocompatibility complex." J Biol Chem **278**(7): 5214-26.

Dorling, A., N. J. Monk, et al. (2000). "HLA-G inhibits the transendothelial migration of human NK cells." Eur J Immunol **30**(2): 586-93.

Dotan, Z. A., A. Dotan, et al. (2000). "Modification in the inherent mode of allelic replication in lymphocytes of patients suffering from renal cell carcinoma: a novel genetic alteration associated with malignancy." Genes Chromosomes Cancer **27**(3): 270-7.

Drouin, R., M. Boutouil, et al. (1997). "DNA replication asynchrony between the paternal and maternal alleles of imprinted genes does not straddle the R/G transition." Chromosoma **106**(6): 405-11.

Drouin, R., G. P. Holmquist, et al. (1994). "High-resolution replication bands compared with morphologic G- and R-bands." Adv Hum Genet **22**: 47-115.

Dutrillaux, B., J. Couturier, et al. (1976). "Sequence of DNA replication in 277 R- and Q-bands of human chromosomes using a BrdU treatment." Chromosoma **58**(1): 51-61.

Edenberg, H. J. and J. A. Huberman (1975). "Eukaryotic chromosome replication." Annu Rev Genet **9**: 245-84.

Eils, R., S. Dietzel, et al. (1996). "Three-dimensional reconstruction of painted human interphase chromosomes: active and inactive X chromosome territories have similar volumes but differ in shape and surface structure." J Cell Biol **135**(6 Pt 1): 1427-40.

Eissner, G., W. Kolch, et al. (2004). "Ligands working as receptors: reverse signaling by members of the TNF superfamily enhance the plasticity of the immune system." Cytokine Growth Factor Rev **15**(5): 353-66.

Ensminger, A. W. and A. Chess (2004). "Coordinated replication timing of monoallelically expressed genes along human autosomes." Hum Mol Genet **13**(6): 651-8.

Epner, E., W. C. Forrester, et al. (1988). "Asynchronous DNA replication within the human beta-globin gene locus." Proc Natl Acad Sci U S A **85**(21): 8081-5.

Eyre-Walker, A. (1992). "Evidence that both G + C rich and G + C poor isochores are replicated early and late in the cell cycle." Nucleic Acids Res **20**(7): 1497-501.

Eyre-Walker, A. and L. D. Hurst (2001). "The evolution of isochores." Nat Rev Genet **2**(7): 549-55.

Felsenfeld, G. and M. Groudine (2003). "Controlling the double helix." Nature **421**(6921): 448-53.

Finch, J. T. and A. Klug (1976). "Solenoidal model for superstructure in chromatin." Proc Natl Acad Sci U S A **73**(6): 1897-901.

Forsburg, S. L. (2004). "Eukaryotic MCM proteins: beyond replication initiation." Microbiol Mol Biol Rev **68**(1): 109-31.

Francke, U. (1994). "Digitized and differentially shaded human chromosome ideograms for genomic applications." Cytogenet Cell Genet **65**(3): 206-18.

Fukagawa, T., Y. Nakamura, et al. (1996). "Human pseudoautosomal boundary-like sequences: expression and involvement in evolutionary formation of the present-day pseudoautosomal boundary of human sex chromosomes." Hum Mol Genet **5**(1): 23-32.

Fukagawa, T., K. Sugaya, et al. (1995). "A boundary of long-range G + C% mosaic domains in the human MHC locus: pseudoautosomal boundary-like sequence exists near the boundary." Genomics **25**(1): 184-91.

Gall, J. G. and H. G. Callan (1962). "H3 uridine incorporation in lampbrush chromosomes." Proc Natl Acad Sci U S A **48**: 562-70.

Gartler, S. M., L. Goldstein, et al. (1999). "The timing of XIST replication: dominance of the domain." Hum Mol Genet **8**(6): 1085-9.

Gasser, S. M., T. Laroche, et al. (1986). "Metaphase chromosome structure. Involvement of topoisomerase II." J Mol Biol **188**(4): 613-29.

Geirsson, A., I. Paliwal, et al. (2003). "Class II transactivator promoter activity is suppressed through regulation by a trophoblast noncoding RNA." Transplantation **76**(2): 387-94.

Giacca, M., C. Pelizon, et al. (1997). "Mapping replication origins by quantifying relative abundance of nascent DNA strands using competitive polymerase chain reaction." Methods **13**(3): 301-12.
Giacca, M., L. Zentilin, et al. (1994). "Fine mapping of a replication origin of human DNA." Proc Natl Acad Sci U S A **91**(15): 7119-23.
Gilbert, D. M. (2002). "Replication timing and transcriptional control: beyond cause and effect." Curr Opin Cell Biol **14**(3): 377-83.
Gomez, M. and N. Brockdorff (2004). "Heterochromatin on the inactive X chromosome delays replication timing without affecting origin usage." Proc Natl Acad Sci U S A **101**(18): 6923-8.
Gullo, C. A. and G. Teoh (2004). "Heat shock proteins: to present or not, that is the question." Immunol Lett **94**(1-2): 1-10.
Han, J. S. and J. D. Boeke (2005). "LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression?" Bioessays **27**(8): 775-84.
Handeli, S., A. Klar, et al. (1989). "Mapping replication units in animal cells." Cell **57**(6): 909-20.
Hansen, R. S., T. K. Canfield, et al. (1995). "Reverse replication timing for the XIST gene in human fibroblasts." Hum Mol Genet **4**(5): 813-20.
Hansen, R. S., T. K. Canfield, et al. (1993). "Association of fragile X syndrome with delayed replication of the FMR1 gene." Cell **73**(7): 1403-9.
Hattori, M., A. Fujiyama, et al. (2000). "The DNA sequence of human chromosome 21." Nature **405**(6784): 311-9.
Holmquist, G., M. Gray, et al. (1982). "Characterization of Giemsa dark- and light-band DNA." Cell **31**(1): 121-9.
Holmquist, G. P. and T. Ashley (2006). "Chromosome organization and chromatin modification: influence on genome function and evolution." Cytogenet Genome Res **114**(2): 96-125.
Honey, K. and A. Y. Rudensky (2003). "Lysosomal cysteine proteases regulate antigen presentation." Nat Rev Immunol **3**(6): 472-82.
Horton, R., L. Wilming, et al. (2004). "Gene map of the extended human MHC." Nat Rev Genet **5**(12): 889-99.
Hozak, P., A. B. Hassan, et al. (1993). "Visualization of replication factories attached to nucleoskeleton." Cell **73**(2): 361-73.
Huberman, J. A. (1997). "Mapping replication origins, pause sites, and termini by neutral/alkaline two-dimensional gel electrophoresis." Methods **13**(3): 247-57.
Huberman, J. A. and A. D. Riggs (1968). "On the mechanism of DNA replication in mammalian chromosomes." J Mol Biol **32**(2): 327-41.
Hughes, E. A. and P. Cresswell (1998). "The thiol oxidoreductase ERp57 is a component of the MHC class I peptide-loading complex." Curr Biol **8**(12): 709-12.
Hyrien, O., K. Marheineke, et al. (2003). "Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem." Bioessays **25**(2): 116-25.
Hyrien, O., C. Maric, et al. (1995). "Transition in specification of embryonic metazoan DNA replication origins." Science **270**(5238): 994-7.
Hyrien, O. and M. Mechali (1993). "Chromosomal replication initiates and terminates at random sequences but at regular intervals in the ribosomal DNA of Xenopus early embryos." Embo J **12**(12): 4511-20.
Ishitani, A., N. Sageshima, et al. (2006). "The involvement of HLA-E and -F in pregnancy." J Reprod Immunol **69**(2): 101-13.

Ishitani, A., N. Sageshima, et al. (2003). "Protein expression and peptide binding suggest unique and interacting functional roles for HLA-E, F, and G in maternal-placental immune recognition." J Immunol **171**(3): 1376-84.

Jackson, D. A. and A. Pombo (1998). "Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells." J Cell Biol **140**(6): 1285-95.

Jacob, F. and S. Brenner (1963). "[On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon.]." C R Hebd Seances Acad Sci **256**: 298-300.

Jeffreys, A. J., J. K. Holloway, et al. (2004). "Meiotic recombination hot spots and human DNA diversity." Philos Trans R Soc Lond B Biol Sci **359**(1441): 141-52.

Jeffreys, A. J., L. Kauppi, et al. (2001). "Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex." Nat Genet **29**(2): 217-22.

Jeon, Y., S. Bekiranov, et al. (2005). "Temporal profile of replication of human chromosomes." Proc Natl Acad Sci U S A **102**(18): 6419-24.

Johnson, A. and M. O'Donnell (2005). "Cellular DNA replicases: components and dynamics at the replication fork." Annu Rev Biochem **74**: 283-315.

Jordan, W. C. and M. W. Bruford (1998). "New perspectives on mate choice and the MHC." Heredity **81 (Pt 3)**: 239-45.

Kagotani, K., S. Takebayashi, et al. (2002). "Replication timing properties within the mouse distal chromosome 7 imprinting cluster." Biosci Biotechnol Biochem **66**(5): 1046-51.

Kamimura, Y., H. Masumoto, et al. (1998). "Sld2, which interacts with Dpb11 in Saccharomyces cerevisiae, is required for chromosomal DNA replication." Mol Cell Biol **18**(10): 6102-9.

Kamimura, Y., Y. S. Tak, et al. (2001). "Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in Saccharomyces cerevisiae." Embo J **20**(8): 2097-107.

Kao, H. I. and R. A. Bambara (2003). "The protein components and mechanism of eukaryotic Okazaki fragment maturation." Crit Rev Biochem Mol Biol **38**(5): 433-52.

Kauppi, L., A. J. Jeffreys, et al. (2004). "Where the crossovers are: recombination distributions in mammals." Nat Rev Genet **5**(6): 413-24.

Kauppi, L., M. P. Stumpf, et al. (2005). "Localized breakdown in linkage disequilibrium does not always predict sperm crossover hot spots in the human MHC class II region." Genomics **86**(1): 13-24.

Kawame, H., S. M. Gartler, et al. (1995). "Allele-specific replication timing in imprinted domains: absence of asynchrony at several loci." Hum Mol Genet **4**(12): 2287-93.

Keller, C., E. M. Ladenburger, et al. (2002). "The origin recognition complex marks a replication origin in the human TOP1 gene promoter." J Biol Chem **277**(35): 31430-40.

Kelley, J., L. Walter, et al. (2005). "Comparative genomics of major histocompatibility complexes." Immunogenetics **56**(10): 683-95.

Kelly, A., S. H. Powis, et al. (1992). "Assembly and function of the two ABC transporter proteins encoded in the human major histocompatibility complex." Nature **355**(6361): 641-4.

Kemp, M. G., M. Ghosh, et al. (2005). "The histone deacetylase inhibitor trichostatin A alters the pattern of DNA replication origin activity in human cells." Nucleic Acids Res **33**(1): 325-36.
Kerem, B. S., R. Goitein, et al. (1984). "Mapping of DNAase I sensitive regions on mitotic chromosomes." Cell **38**(2): 493-9.
Kimura, K., O. Cuvier, et al. (2001). "Chromosome condensation by a human condensin complex in Xenopus egg extracts." J Biol Chem **276**(8): 5417-20.
Kitsberg, D., S. Selig, et al. (1993). "Allele-specific replication timing of imprinted gene regions." Nature **364**(6436): 459-63.
Kitsberg, D., S. Selig, et al. (1993). "Replication structure of the human beta-globin gene domain." Nature **366**(6455): 588-90.
Korenberg, J. R. and M. C. Rykowski (1988). "Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands." Cell **53**(3): 391-400.
Kreahling, J. and B. R. Graveley (2004). "The origins and implications of Aluternative splicing." Trends Genet **20**(1): 1-4.
Kreitz, S., M. Ritzi, et al. (2001). "The human origin recognition complex protein 1 dissociates from chromatin during S phase in HeLa cells." J Biol Chem **276**(9): 6337-42.
Krishan, A. (1975). "Rapid flow cytofluorometric analysis of mammalian cell cycle by propidium iodide staining." J Cell Biol **66**(1): 188-93.
Kumar, S., M. Giacca, et al. (1996). "Utilization of the same DNA replication origin by human cells of different derivation." Nucleic Acids Res **24**(17): 3289-94.
Ladenburger, E. M., C. Keller, et al. (2002). "Identification of a binding region for human origin recognition complex proteins 1 and 2 that coincides with an origin of DNA replication." Mol Cell Biol **22**(4): 1036-48.
Lande-Diner, L. and H. Cedar (2005). "Silence of the genes--mechanisms of long-term repression." Nat Rev Genet **6**(8): 648-54.
Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
Landmann, S., A. Muhlethaler-Mottet, et al. (2001). "Maturation of dendritic cells is accompanied by rapid transcriptional silencing of class II transactivator (CIITA) expression." J Exp Med **194**(4): 379-91.
Larsen, C. E. and C. A. Alper (2004). "The genetics of HLA-associated disease." Curr Opin Immunol **16**(5): 660-7.
Lehman, I. R. and L. S. Kaguni (1989). "DNA polymerase alpha." J Biol Chem **264**(8): 4265-8.
Leonhardt, H., H. P. Rahn, et al. (2000). "Dynamics of DNA replication factories in living cells." J Cell Biol **149**(2): 271-80.
Lepin, E. J., J. M. Bastin, et al. (2000). "Functional characterization of HLA-F and binding of HLA-F tetramers to ILT2 and ILT4 receptors." Eur J Immunol **30**(12): 3552-61.
Leuba, S. H., G. Yang, et al. (1994). "Three-dimensional structure of extended chromatin fibers as revealed by tapping-mode scanning force microscopy." Proc Natl Acad Sci U S A **91**(24): 11621-5.
Li, C. J. and M. L. DePamphilis (2002). "Mammalian Orc1 protein is selectively released from chromatin and ubiquitinated during the S-to-M transition in the cell division cycle." Mol Cell Biol **22**(1): 105-16.

Li, L. and L. Zou (2005). "Sensing, signaling, and responding to DNA damage: organization of the checkpoint pathways in mammalian cells." J Cell Biochem **94**(2): 298-306.
Li, Z., A. Menoret, et al. (2002). "Roles of heat-shock proteins in antigen presentation and cross-presentation." Curr Opin Immunol **14**(1): 45-51.
Lichter, P., T. Cremer, et al. (1988). "Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries." Hum Genet **80**(3): 224-34.
Lin, C. M., H. Fu, et al. (2003). "Dynamic alterations of replication timing in mammalian cells." Curr Biol **13**(12): 1019-28.
Liu, G., M. Malott, et al. (2003). "Multiple functional elements comprise a Mammalian chromosomal replicator." Mol Cell Biol **23**(5): 1832-42.
Locksley, R. M., N. Killeen, et al. (2001). "The TNF and TNF receptor superfamilies: integrating mammalian biology." Cell **104**(4): 487-501.
Lu, Z. H., D. B. Sittman, et al. (1998). "Histone H1 reduces the frequency of initiation in Xenopus egg extract by limiting the assembly of prereplication complexes on sperm chromatin." Mol Biol Cell **9**(5): 1163-76.
Luger, K., A. W. Mader, et al. (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-60.
Ma, H., J. Samarabandu, et al. (1998). "Spatial and temporal dynamics of DNA replication sites in mammalian cells." J Cell Biol **143**(6): 1415-25.
MacAlpine, D. M., H. K. Rodriguez, et al. (2004). "Coordination of replication and transcription along a Drosophila chromosome." Genes Dev **18**(24): 3094-105.
Macaya, G., J. P. Thiery, et al. (1976). "An approach to the organization of eukaryotic genomes at a macromolecular level." J Mol Biol **108**(1): 237-54.
Mallya, M., R. D. Campbell, et al. (2002). "Transcriptional analysis of a novel cluster of LY-6 family members in the human and mouse major histocompatibility complex: five genes with many splice forms." Genomics **80**(1): 113-23.
Manders, E. M., J. Stap, et al. (1992). "Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy." J Cell Sci **103 (Pt 3)**: 857-62.
Manuelidis, L. (1985). "Individual interphase chromosome domains revealed by in situ hybridization." Hum Genet **71**(4): 288-93.
Marsden, M. P. and U. K. Laemmli (1979). "Metaphase chromosome structure: evidence for a radial loop model." Cell **17**(4): 849-58.
McGowan, C. H. and P. Russell (2004). "The DNA damage response: sensing and signaling." Curr Opin Cell Biol **16**(6): 629-33.
Melixetian, M., A. Ballabeni, et al. (2004). "Loss of Geminin induces rereplication in the presence of functional p53." J Cell Biol **165**(4): 473-82.
Mendez, J. and B. Stillman (2000). "Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis." Mol Cell Biol **20**(22): 8602-12.
Mendez, J. and B. Stillman (2003). "Perpetuating the double helix: molecular machines at eukaryotic DNA replication origins." Bioessays **25**(12): 1158-67.
Mesner, L. D., X. Li, et al. (2003). "The dihydrofolate reductase origin of replication does not contain any nonredundant genetic elements required for origin activity." Mol Cell Biol **23**(3): 804-14.

Milinski, M., S. Griffiths, et al. (2005). "Mate choice decisions of stickleback females predictably modified by MHC peptide ligands." Proc Natl Acad Sci U S A **102**(12): 4414-8.

Mostoslavsky, R., N. Singh, et al. (2001). "Asynchronous replication and allelic exclusion in the immune system." Nature **414**(6860): 221-5.

Mouchiroud, D., G. D'Onofrio, et al. (1991). "The distribution of genes in the human genome." Gene **100**: 181-7.

Munkel, C., R. Eils, et al. (1999). "Compartmentalization of interphase chromosomes observed in simulation and experiment." J Mol Biol **285**(3): 1053-65.

Muromoto, R., K. Sugiyama, et al. (2004). "Physical and functional interactions between Daxx and DNA methyltransferase 1-associated protein, DMAP1." J Immunol **172**(5): 2985-93.

Nagarajan, U. M., J. Lochamy, et al. (2002). "Class II transactivator is required for maximal expression of HLA-DOB in B cells." J Immunol **168**(4): 1780-6.

Nakamura, H., T. Morita, et al. (1986). "Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleus." Exp Cell Res **165**(2): 291-7.

Nakayasu, H. and R. Berezney (1989). "Mapping replicational sites in the eucaryotic cell nucleus." J Cell Biol **108**(1): 1-11.

Narlikar, G. J., H. Y. Fan, et al. (2002). "Cooperation between complexes that regulate chromatin structure and transcription." Cell **108**(4): 475-87.

Newkirk, H. L., J. H. Knoll, et al. (2005). "Distortion of quantitative genomic and expression hybridization by Cot-1 DNA: mitigation of this effect." Nucleic Acids Res **33**(22): e191.

Nishitani, H., Z. Lygerou, et al. (2004). "Proteolysis of DNA replication licensing factor Cdt1 in S-phase is performed independently of geminin through its N-terminal region." J Biol Chem **279**(29): 30807-16.

Nishitani, H., S. Taraviras, et al. (2001). "The human licensing factor for DNA replication Cdt1 accumulates in G1 and is destabilized after initiation of S-phase." J Biol Chem **276**(48): 44905-11.

Norio, P., S. Kosiyatrakul, et al. (2005). "Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development." Mol Cell **20**(4): 575-87.

O'Keefe, R. T., S. C. Henderson, et al. (1992). "Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences." J Cell Biol **116**(5): 1095-110.

Ohno, S. (1972). "So much "junk" DNA in our genome." Brookhaven Symp Biol **23**: 366-70.

Oliver, J. D., H. L. Roderick, et al. (1999). "ERp57 functions as a subunit of specific complexes formed with the ER lectins calreticulin and calnexin." Mol Biol Cell **10**(8): 2573-82.

Ortmann, B., J. Copeman, et al. (1997). "A critical role for tapasin in the assembly and function of multimeric MHC class I-TAP complexes." Science **277**(5330): 1306-9.

Ouchi, T., S. W. Lee, et al. (2000). "Collaboration of signal transducer and activator of transcription 1 (STAT1) and BRCA1 in differential regulation of IFN-gamma target genes." Proc Natl Acad Sci U S A **97**(10): 5208-13.

Paixao, S., I. N. Colaluca, et al. (2004). "Modular structure of the human lamin B2 replicator." Mol Cell Biol **24**(7): 2958-67.

Pardee, A. B. (1989). "G1 events and regulation of cell proliferation." Science **246**(4930): 603-8.
Parham, P. (1996). "Functions for MHC class I carbohydrates inside and outside the cell." Trends Biochem Sci **21**(11): 427-33.
Paulson, J. R. and U. K. Laemmli (1977). "The structure of histone-depleted metaphase chromosomes." Cell **12**(3): 817-28.
Pavlicek, A., O. Clay, et al. (2002). "Isochore conservation between MHC regions on human chromosome 6 and mouse chromosome 17." FEBS Lett **511**(1-3): 175-7.
Payne, R. and M. R. Rolfs (1958). "Fetomaternal leukocyte incompatibility." J Clin Invest **37**(12): 1756-63.
Payne, R., M. Tripp, et al. (1964). "A New Leukocyte Isoantigen System In Man." Cold Spring Harb Symp Quant Biol **29**: 285-95.
Pedrali-Noy, G., S. Spadari, et al. (1980). "Synchronization of HeLa cell cultures by inhibition of DNA polymerase alpha with aphidicolin." Nucleic Acids Res **8**(2): 377-87.
Pende, D., S. Parolini, et al. (1999). "Identification and molecular characterization of NKp30, a novel triggering receptor involved in natural cytotoxicity mediated by human natural killer cells." J Exp Med **190**(10): 1505-16.
Perry, P., S. Sauer, et al. (2004). "A dynamic switch in the replication timing of key regulator genes in embryonic stem cells upon neural induction." Cell Cycle **3**(12): 1645-50.
Piskurich, J. F., K. I. Lin, et al. (2000). "BLIMP-I mediates extinction of major histocompatibility class II transactivator expression in plasma cells." Nat Immunol **1**(6): 526-32.
Puck, T. T., S. J. Cieciura, et al. (1958). "GENETICS OF SOMATIC MAMMALIAN CELLS: III. LONG-TERM CULTIVATION OF EUPLOID CELLS FROM HUMAN AND ANIMAL SUBJECTS." J. Exp. Med. **108**(6): 945-956.
Raghuraman, M. K., E. A. Winzeler, et al. (2001). "Replication dynamics of the yeast genome." Science **294**(5540): 115-21.
Rapaport, F. T. and J. Dausset (1968). "Tissue typing in human transplantation." J Am Med Womens Assoc **23**(11): 1029-40.
Razin, A. and H. Cedar (1994). "DNA methylation and genomic imprinting." Cell **77**(4): 473-6.
Rein, T., T. Kobayashi, et al. (1999). "DNA methylation at mammalian replication origins." J Biol Chem **274**(36): 25792-800.
Reish, O., A. Orlovski, et al. (2003). "Modified allelic replication in lymphocytes of patients with neurofibromatosis type 1." Cancer Genet Cytogenet **143**(2): 133-9.
Remus, D., E. L. Beall, et al. (2004). "DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding." Embo J **23**(4): 897-907.
Riggs, A. D. and G. P. Pfeifer (1992). "X-chromosome inactivation and cell memory." Trends Genet **8**(5): 169-74.
Rollinger-Holzinger, I., B. Eibl, et al. (2000). "LST1: a gene with extensive alternative splicing and immunomodulatory function." J Immunol **164**(6): 3169-76.
Saccone, S. and G. Bernardi (2001). "Human chromosomal banding by in situ hybridization of isochores." Methods Cell Sci **23**(1-3): 7-15.
Saccone, S., S. Caccio, et al. (1996). "Identification of the gene-richest bands in human chromosomes." Gene **174**(1): 85-94.

Sadoni, N., S. Langer, et al. (1999). "Nuclear organization of mammalian genomes. Polar chromosome territories build up functionally distinct higher order compartments." J Cell Biol **146**(6): 1211-26.

Saitoh, Y. and U. K. Laemmli (1994). "Metaphase chromosome structure: bands arise from a differential folding path of the highly AT-rich scaffold." Cell **76**(4): 609-22.

Salomoni, P. and A. F. Khelifi (2006). "Daxx: death or survival protein?" Trends Cell Biol **16**(2): 97-104.

Santocanale, C. and J. F. Diffley (1998). "A Mec1- and Rad53-dependent checkpoint controls late-firing origins of DNA replication." Nature **395**(6702): 615-8.

Santocanale, C., K. Sharma, et al. (1999). "Activation of dormant origins of DNA replication in budding yeast." Genes Dev **13**(18): 2360-4.

Schaarschmidt, D., J. Baltin, et al. (2004). "An episomal mammalian replicon: sequence-independent binding of the origin recognition complex." Embo J **23**(1): 191-201.

Schardin, M., T. Cremer, et al. (1985). "Specific staining of human chromosomes in Chinese hamster x man hybrid cell lines demonstrates interphase chromosome territories." Hum Genet **71**(4): 281-7.

Schmidt, M. and B. R. Migeon (1990). "Asynchronous replication of homologous loci on human active and inactive X chromosomes." Proc Natl Acad Sci U S A **87**(10): 3685-9.

Schubeler, D., D. Scalzo, et al. (2002). "Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing." Nat Genet **32**(3): 438-42.

Selig, S., K. Okumura, et al. (1992). "Delineation of DNA replication time zones by fluorescence in situ hybridization." Embo J **11**(3): 1217-25.

Senger, G., J. Ragoussis, et al. (1993). "Fine mapping of the human MHC class II region within chromosome band 6p21 and evaluation of probe ordering using interphase fluorescence in situ hybridization." Cytogenet Cell Genet **64**(1): 49-53.

Shechter, D. and J. Gautier (2004). "MCM proteins and checkpoint kinases get together at the fork." Proc Natl Acad Sci U S A **101**(30): 10845-6.

Shen, L., J. Guo, et al. (2006). "Distinct domains for anti- and pro-apoptotic activities of IEX-1." J Biol Chem.

Sherr, C. J. (1994). "G1 phase progression: cycling on cue." Cell **79**(4): 551-5.

Sherr, C. J. and J. M. Roberts (1999). "CDK inhibitors: positive and negative regulators of G1-phase progression." Genes Dev **13**(12): 1501-12.

Shmueli, O., S. Horn-Saban, et al. (2003). "GeneNote: whole genome expression profiles in normal human tissues." C R Biol **326**(10-11): 1067-72.

Sim, R. B. and S. A. Tsiftsoglou (2004). "Proteases of the complement system." Biochem Soc Trans **32**(Pt 1): 21-7.

Simon, I., T. Tenzen, et al. (2001). "Developmental regulation of DNA replication timing at the human beta globin locus." Embo J **20**(21): 6150-7.

Singh, N., F. A. Ebrahimi, et al. (2003). "Coordination of the random asynchronous replication of autosomal loci." Nat Genet **33**(3): 339-41.

Smith, R. A., P. J. Ho, et al. (1998). "Recombination breakpoints in the human beta-globin gene cluster." Blood **92**(11): 4415-21.

Smith, Z. E. and D. R. Higgs (1999). "The pattern of replication at a human telomeric region (16p13.3): its relationship to chromosome structure and gene expression." Hum Mol Genet **8**(8): 1373-86.

Smyth, G. K. (2005). limma: Linear Models for Microarray data. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. R. C. Gentleman, V.; Huber, W.; Irizarry, R.; Dudoit, S., Springer: 397-420.

Srivastava, P. (2002). "Interaction of heat shock proteins with peptides and antigen presenting cells: chaperoning of the innate and adaptive immune responses." Annu Rev Immunol **20**: 395-425.

Srivastava, P. (2002). "Roles of heat-shock proteins in innate and adaptive immunity." Nat Rev Immunol **2**(3): 185-94.

Stark, G. R., I. M. Kerr, et al. (1998). "How cells respond to interferons." Annu Rev Biochem **67**: 227-64.

Stephanou, A., D. A. Isenberg, et al. (1999). "Signal transducer and activator of transcription-1 and heat shock factor-1 interact and activate the transcription of the Hsp-70 and Hsp-90beta gene promoters." J Biol Chem **274**(3): 1723-8.

Stephanou, A. and D. S. Latchman (1999). "Transcriptional regulation of the heat shock protein genes by STAT family transcription factors." Gene Expr **7**(4-6): 311-9.

Stephens, H. A. (2001). "MICA and MICB genes: can the enigma of their polymorphism be resolved?" Trends Immunol **22**(7): 378-85.

Stephens, R., R. Horton, et al. (1999). "Gene organisation, sequence variation and isochore structure at the centromeric boundary of the human MHC." J Mol Biol **291**(4): 789-99.

Stern, L. J., J. H. Brown, et al. (1994). "Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide." Nature **368**(6468): 215-21.

Stewart, C. A., R. Horton, et al. (2004). "Complete MHC haplotype sequencing for common disease gene mapping." Genome Res **14**(6): 1176-87.

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature **403**(6765): 41-5.

Strehl, S., J. M. LaSalle, et al. (1997). "High-resolution analysis of DNA replication domain organization across an R/G-band boundary." Mol Cell Biol **17**(10): 6157-66.

Sugimoto, N., Y. Tatsumi, et al. (2004). "Cdt1 phosphorylation by cyclin A-dependent kinases negatively regulates its function without affecting geminin binding." J Biol Chem **279**(19): 19691-7.

Sun, Y., R. T. Wyatt, et al. (2001). "Expression and replication timing patterns of wildtype and translocated BCL2 genes." Genomics **73**(2): 161-70.

Svetlova, E. Y., S. V. Razin, et al. (2001). "Mammalian recombination hot spot in a DNA loop anchorage region: A model for the study of common fragile sites." J Cell Biochem **81**(S36): 170-178.

Szanto, A., V. Narkar, et al. (2004). "Retinoid X receptors: X-ploring their (patho)physiological functions." Cell Death Differ **11 Suppl 2**: S126-43.

Taddeo, B., A. Esclatine, et al. (2003). "The stress-inducible immediate-early responsive gene IEX-1 is activated in cells infected with herpes simplex virus 1, but several viral mechanisms, including 3' degradation of its RNA, preclude expression of the gene." J Virol **77**(11): 6178-87.

Taira, T., S. M. Iguchi-Ariga, et al. (1994). "A novel DNA replication origin identified in the human heat shock protein 70 gene promoter." Mol Cell Biol **14**(9): 6386-97.

Takayama, Y., Y. Kamimura, et al. (2003). "GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast." Genes Dev **17**(9): 1153-65.
Takebayashi, S., K. Sugimura, et al. (2005). "Regulation of replication at the R/G chromosomal band boundary and pericentromeric heterochromatin of mammalian cells." Exp Cell Res **304**(1): 162-74.
Tanaka, K. (1998). "Molecular biology of the proteasome." Biochem Biophys Res Commun **247**(3): 537-41.
Tao, L., Z. Dong, et al. (2000). "Major DNA replication initiation sites in the c-myc locus in human cells." J Cell Biochem **78**(3): 442-57.
Taylor, J. (2005). "Clues to function in gene deserts." Trends Biotechnol **23**(6): 269-71.
Tenzen, T., T. Yamagata, et al. (1997). "Precise switching of DNA replication timing in the GC content transition area in the human major histocompatibility complex." Mol Cell Biol **17**(7): 4043-50.
Thiery, J. P., G. Macaya, et al. (1976). "An analysis of eukaryotic genomes by density gradient centrifugation." J Mol Biol **108**(1): 219-35.
Ting, A. and P. I. Terasaki (1974). "Influence of lymphocyte-dependent antibodies on human kidney transplants." Transplantation **18**(4): 371-3.
Ting, J. P. and J. Trowsdale (2002). "Genetic control of MHC class II expression." Cell **109 Suppl**: S21-33.
Toledo, F., B. Baron, et al. (1998). "oriGNAI3: a narrow zone of preferential replication initiation in mammalian cells identified by 2D gel and competitive PCR replicon mapping techniques." Nucleic Acids Res **26**(10): 2313-21.
Torchia, B. S., L. M. Call, et al. (1994). "DNA replication analysis of FMR1, XIST, and factor 8C loci by FISH shows nontranscribed X-linked genes replicate late." Am J Hum Genet **55**(1): 96-104.
Torchia, B. S. and B. R. Migeon (1995). "The XIST locus replicates late on the active X, and earlier on the inactive X based on FISH DNA replication analysis of somatic cell hybrids." Somat Cell Mol Genet **21**(5): 327-33.
Traherne, J. A., R. Horton, et al. (2006). "Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history." PLoS Genet **2**(1): e9.
Tribioli, C., G. Biamonti, et al. (1987). "Characterization of human DNA sequences synthesized at the onset of S-phase." Nucleic Acids Res **15**(24): 10211-32.
Trowsdale, J. (2002). "The gentle art of gene arrangement: the meaning of gene clusters." Genome Biol **3**(3): COMMENT2002.
Trowsdale, J. and P. Parham (2004). "Mini-review: defense strategies and immunity-related genes." Eur J Immunol **34**(1): 7-17.
van Dierendonck, J. H., R. Keyzer, et al. (1989). "Subdivision of S-phase by analysis of nuclear 5-bromodeoxyuridine staining patterns." Cytometry **10**(2): 143-50.
van Drunen, C. M., R. G. Sewalt, et al. (1999). "A bipartite sequence element associated with matrix/scaffold attachment regions." Nucleic Acids Res **27**(14): 2924-30.
Van Rood, J. J., J. G. Eernisse, et al. (1958). "Leucocyte antibodies in sera from pregnant women." Nature **181**(4625): 1735-6.
Vashee, S., C. Cvetic, et al. (2003). "Sequence-independent DNA binding and replication initiation by the human origin recognition complex." Genes Dev **17**(15): 1894-908.

Vassilev, L. and E. M. Johnson (1990). "An initiation zone of chromosomal DNA replication located upstream of the c-myc gene in proliferating HeLa cells." Mol Cell Biol **10**(9): 4899-904.

Vaughn, J. P., P. A. Dijkwel, et al. (1990). "Replication initiates in a broad zone in the amplified CHO dihydrofolate reductase domain." Cell **61**(6): 1075-87.

Verbovaia, L. V. and S. V. Razin (1997). "Mapping of replication origins and termination sites in the Duchenne muscular dystrophy gene." Genomics **45**(1): 24-30.

Verma, R. S. and A. Babu (1995). Human chromosomes: principles and techniques. New York; London, McGraw-Hill.

Volpi, E. V., E. Chevret, et al. (2000). "Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei." J Cell Sci **113 (Pt 9)**: 1565-76.

Wall, J. D., L. A. Frisse, et al. (2003). "Comparative linkage-disequilibrium analysis of the beta-globin hotspot in primates." Am J Hum Genet **73**(6): 1330-40.

Wang, L., J. Darling, et al. (1999). "Allele-specific late replication and fragility of the most active common fragile site, FRA3B." Hum Mol Genet **8**(3): 431-7.

Wang, S., P. A. Dijkwel, et al. (1998). "Lagging-strand, early-labelling, and two-dimensional gel assays suggest multiple potential initiation sites in the Chinese hamster dihydrofolate reductase origin." Mol Cell Biol **18**(1): 39-50.

Watanabe, Y., A. Fujiyama, et al. (2002). "Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: disease-related genes in timing-switch regions." Hum Mol Genet **11**(1): 13-21.

Watanabe, Y., T. Tenzen, et al. (2000). "Replication timing of the human X-inactivation center (XIC) region: correlation with chromosome bands." Gene **252**(1-2): 163-72.

Weinberg, R. A. (1995). "The retinoblastoma protein and cell cycle control." Cell **81**(3): 323-30.

White, E. J., O. Emanuelsson, et al. (2004). "DNA replication-timing analysis of human chromosome 22 at high resolution and different developmental states." Proc Natl Acad Sci U S A **101**(51): 17771-6.

Williams, B. A., R. M. Gwirtz, et al. (2004). "Genomic DNA as a cohybridization standard for mammalian microarray measurements." Nucleic Acids Res **32**(10): e81.

Wohlschlegel, J. A., B. T. Dwyer, et al. (2000). "Inhibition of eukaryotic DNA replication by geminin binding to Cdt1." Science **290**(5500): 2309-12.

Wold, M. S. (1997). "Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism." Annu Rev Biochem **66**: 61-92.

Woodcock, C. L., S. A. Grigoryev, et al. (1993). "A chromatin folding model that incorporates linker variability generates fibers resembling the native structures." Proc Natl Acad Sci U S A **90**(19): 9021-5.

Woodfine, K., D. M. Beare, et al. (2005). "Replication timing of human chromosome 6." Cell Cycle **4**(1): 172-6.

Woodfine, K., N. P. Carter, et al. (2005). "Investigating chromosome organization with genomic microarrays." Chromosome Res **13**(3): 249-57.

Woodfine, K., H. Fiegler, et al. (2004). "Replication timing of the human genome." Hum Mol Genet **13**(2): 191-202.

Wu, J., V. Groh, et al. (2002). "T cell antigen receptor engagement and specificity in the recognition of stress-inducible MHC class I-related chains by human epithelial gamma delta T cells." J Immunol **169**(3): 1236-40.

Wu, J. R. and D. M. Gilbert (1996). "A distinct G1 step required to specify the Chinese hamster DHFR replication origin." Science **271**(5253): 1270-2.

Xiong, Z., W. Tsark, et al. (1998). "Differential replication timing of X-linked genes measured by a novel method using single-nucleotide primer extension." Nucleic Acids Res **26**(2): 684-6.

Yunis, J. J. (1981). "Mid-prophase human chromosomes. The attainment of 2000 bands." Hum Genet **56**(3): 293-8.

Yunis, J. J., L. Roldan, et al. (1971). "Staining of satellite DNA in metaphase chromosomes." Nature **231**(5304): 532-3.

Yurov, Y. B. and N. A. Liapunova (1977). "The units of DNA replication in the mammalian chromosomes: evidence for a large size of replication units." Chromosoma **60**(3): 253-67.

Zhang, C. T. and R. Zhang (2003). "An isochore map of the human genome based on the Z curve method." Gene **317**(1-2): 127-35.

Zhang, J., F. Xu, et al. (2002). "Establishment of transcriptional competence in early and late S phase." Nature **420**(6912): 198-202.

Zhang, J. J., U. Vinkemeier, et al. (1996). "Two contact regions between Stat1 and CBP/p300 in interferon gamma signaling." Proc Natl Acad Sci U S A **93**(26): 15092-6.

Zhang, J. J., Y. Zhao, et al. (1998). "Ser727-dependent recruitment of MCM5 by Stat1alpha in IFN-gamma-induced transcriptional activation." Embo J **17**(23): 6963-71.

Zirbel, R. M., U. R. Mathieu, et al. (1993). "Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries." Chromosome Res **1**(2): 93-106.

Zoubak, S., O. Clay, et al. (1996). "The gene distribution of the human genome." Gene **174**(1): 95-102.